

46
1
中國協和醫科大學
中國醫學科學院

博士研究生學位論文

中國協和醫科大學研究生院

河豚鱼多药耐药基因的克隆

河豚鱼多药耐药基因组全长序列和邻座基因的分析

博士研究生: 刘 勇

导 师: 杨焕明 (基础所)

方福德 (基础所)

杨纯正 (血液所)

学科专业: 分子遗传学

入学时间: 一九九六年九月

所在所院: 中国协和医科大学基础医学院

一九九八年六月



*Apologize for always going home at midnight
In order to finish this thesis in two years*

To my wife

目 录

中文摘要.....	1
英文摘要.....	2
前 言.....	3
第一部分.....	10
试剂与材料.....	10
方 法.....	12
结 果.....	22
讨 论.....	37
第二部分.....	42
方法与结果.....	42
讨 论.....	193
文 献.....	211
综 述 一.....	212
综 述 二.....	232
附 录.....	252
致 谢.....	260

摘 要

人的多药耐药基因家族是由MDR1和MDR3组成,二者位于7q21.1,总跨度约230kb,其中MDR1>120kb,MDR3为74kb。其表达产物为一跨膜蛋白p-glycoprotein,通过ATP提供能量,将进入细胞内的多种物质泵出细胞外,从而使细胞对这些物质产生耐药性,多药耐药基因的功能决定了它在肿瘤化疗中的位置而使其成为基因结构、功能和表达、调控研究中的热点。本文采用人的多药耐药基因的全长cDNA为探针,通过与河豚鱼cosmid文库的杂交获得候选克隆,再通过反复杂交验证最终获得9个多药耐药基因的阳性克隆,通过ExoIII系统和shotgun法获得cosmid 124A22的全长序列。采用生物信息学的多种软件系统,对全长序列进行分析、预测。结果表明:河豚鱼基因组中含有至少三个多药耐药基因,以头尾方式相连,在cosmid 124A22中含有两个阅读框架完整的多药耐药基因,全长序列分别为10.8kb和10.2kb,只有人的多药耐药基因的1/7-1/12,但二者的cDNA(exon4-exon28)分别为3732bp和3699bp,与人的3723bp(exon4-exon28)无论在核苷酸水平还是氨基酸水平都表现出较高的同源性,氨基酸水平的同源性分别为75%和76%,其结构也高度保守,包括跨膜结构,功能保守区即核苷酸结合位点高度保守。在这40kb的序列中共包括6.5个基因,基因的平均长度为6kb。这些结果都表明河豚鱼作为一种新型模式生物可以在比较基因组学、基因的结构功能和表达调控等方面发挥积极、重要的作用。

关键词: 河豚鱼 多药耐药基因 杂交 序列

Abstract

The human MDR (multil-drug resistance) gene family is composed of two members (MDR1, MDR3) which are clustered on the chromosome band 7q21.1 and span a region of about 230 kb. A human MDR1 full length cDNA probe was used to screen a pufferfish cosmid library and nine clones, which would contain the whole MDR gene family in the pufferfish genome, were finally identified. One of the cosmid clones was sequenced. Sequence analysis revealed that there would be at least three MDR homologues in the corresponding region. The biggest MDR gene is of about 10.8 kb, only approximately 1/7 to 1/12 that of the human gene, but its transcript, approximately 3.7 kb long from exon4 to exon28, is in similar size to that of the human. Its gene organization is highly conserved, especially the 12 trans-membranous domains which have been found in the MDR gene family of many organisms. The ATP-binding motif of the gene is significantly homologous. The other two MDR homologues are similar in structures but even smaller. There are total 6.5 genes in the cosmid, with full length of 40kb, and the average size of genes is 6kb. Our preliminary data indicate that pufferfish is an ideal model for the studies of vertebrate genomes and potential for identification of human genes because of its compactness and the conserved gene order in the same evolutionary blocks.

Keywords: Pufferfish Multi-drug resistance Hybridiazation
Sequence

前 言

河豚鱼作为模式生物的作用和意义

九十年代开展的人类基因组计划是本世纪最伟大的三大计划之一。该计划的近期目标是完成四张图——遗传图、物理图、序列图和转录图的绘制，它的最终目的，也是一个无截止日期的目标是找出人类基因组所有的基因以及这些基因的结构、功能和表达、调控。为了达到这些目的，一个特殊的辅助工具由此而产生——模式生物。现已被人类基因组计划正式接受的模式生物有：酵母、线虫、果蝇和小鼠。它们所起的作用是其作为非模式生物的工具所无法替代的。酵母作为最简单的真核生物，一直是作为研究真核生物基因（主要是单基因）的结构、功能和表达、调控的有效模型；线虫由于其具有躯体透明、结构简单（为简单的多细胞生物）而一直成为研究细胞分化和组织发育的最理想模型；果蝇作为经典遗传学的“功臣”，不仅仅与遗传连锁率的确立有关，由于其特殊的染色体结构，又使其与细胞遗传学中多线染色体的发现以及分子遗传学中同源盒的发现密不可分；小鼠与人同属哺乳类，由于其繁殖周期短、经济易得、许多生理和病理现象与人类类似，使其一直成为研究遗传、病理、生理、药理的最佳模型。其中酵母、线虫、果蝇的基因组大小分别为 14Mb(megabase)、100Mb 和 165Mb，远较人类基因组 3000Mb 小得多，对它们的基因组作图，在很大程度上起到人类基因组作图的参照系的作用，但它们又有其不足之处：酵母、线虫、果蝇与人类的进化距离相距甚远，其基因组无论是结构还是功能、更多的是基因板块的保守性与人类基因组差异太大；小鼠与人的进化距离很近，但其基因组与人类的一样复杂。由于人类基因组 94%的区域都为基因间序列和内含子序列，这些序列尤其是其中的大量的重复序列的存在给实际工作造成了诸多困难：1. 使得基因组的各亚克隆之间很难组成“克隆重叠群”，同时，在完成序列图的过程中，重复序列的存在造成了大量的无效劳动和财力浪费；2. 单拷贝序列很难在整个基因组中准确定向定位，分离表达序列工作（如致病基因克隆）显得十分艰难，犹如大海捞针；3. 在研究基因的表达、调控的过程中，一方面会由于庞大的内含子序列而使转基因工作（包括细胞转染和转基因动物）难度加大甚至不可为，另一方面会因缺少存在于内含子中的一些增强子及其一些尚未知晓的内含子特殊功能的作用，而使单纯利用 cDNA 来进行研究的过程“失真”，更谈不上进行基因组功能的研究了。

这样，在完善现有的一些技术、设备之外，寻找一种新型模式生物成为人类基因组计划中迫在眉睫的任务之一：这种模式生物是人类基因组的“作战沙盘”，

该“沙盘”具有以下特征：①各独立单位基本存在，且去粗存精，性质基本不变；②各单位之间的相对关系（包括位置、作用）基本不变。组成基因组的各单位就是散布于人的 24 条染色体的近 100000 个基因，基因内的独立单位就是一个个外显子，其性质主要指基因的功能，各单位之间的相对关系就是指基因之间的调控及空间关系。

1968 年，Ralph.Hinegardner 分析了 100 多种硬骨鱼的细胞 DNA 总量，发现豚科动物，包括河豚鱼含有较小的基因组；80 年代，Pizon 等又对其进行了验证，结果显示河豚鱼的基因组含量为 0.4-0.5pg（单倍体），约相当于 400-500Mb。

河豚鱼属于脊索动物门，硬骨鱼纲，辐鳍亚纲，全骨总目，豚形目，豚亚目，豚科，东方豚属，与人的进化距离约 450 万年。常见的有虫纹东方豚（*Fugu Vermicularis*）等。

1993 年，剑桥大学的 Brenner 系统的研究了河豚鱼基因组的特点：

1. 通过 Southern 杂交发现在平均 24625 个克隆中可含有任何一个基因，按每个克隆平均含有 16.5Kb 的插入片段计算，河豚鱼基因组大小为 $24625 \times 16.5\text{Kb} = 404\text{Mb}$ ；

2. 通过河豚鱼 596 个克隆测序结果分析，发现所测的 127831 个碱基中有 1011 个碱基为编码序列（通过与 Genebank 比较得知，这又称为电子杂交）。由此推断河豚鱼基因组中编码序列的比例是 $1011/127831 = 0.791\%$ ，而人的比例为 0.103%，这说明河豚鱼基因组中基因密度为人的 $0.791/0.103 = 7.68$ 倍。如果河豚鱼与人具有相同的基因数，那么河豚鱼的基因组大小即为 $3000\text{Mb}/7.68 = 390\text{Mb}$ ，这都与 Chuou-Houki 等验证的硬骨鱼类的基因组为 380Mb 相一致，一般现将河豚鱼的基因组大小定为 400Mb。在 127831 个碱基中，重复序列为 9762bp，占 7.6%，且都成串排列，缺少 Alu 或 LINE 那样在人类基因组中散在分布的重复序列。这就是说其中 92.4% 的基因组为单一序列，而人类只有 40% 为单一序列；

3. 人类基因组中有 3% 的序列被转录，按每个转录本平均为 1000bp 来算，人类共有 $3\% \times 3 \times 10^9 \text{bp} / 1000\text{bp} = 90000$ 个基因；根据河豚鱼文库某一 cosmid 全长序列（约 40Kb）中含有 5773bp 的可编码碱基来计算，河豚鱼基因组约含 $5773\text{bp} / 40\text{Kb} \times (4 \times 10^8) / 1000\text{bp} = 60000$ 个基因，虽然计算方法不尽准确，只是粗略估计，但总的趋势说明河豚鱼与人有着基本相同的基因数目；

4. 河豚鱼基因组的 G+C% 为 44.2%，没有明显的 G 带（G、C 稀疏区）和 R 带（G、C 富含区），这也说明河豚鱼的所有基因在基因组中均匀分布。

对基因组的研究发现：

1. 基因间的序列短，基因与基因之间呈高度浓缩状，如 Trower 等发现一个 40Kb 的 cosmid 中含有 8 个基因，其中 3 个基因的总长为 12.4Kb，而与其对应的人 14q24.3 的同源基因组却跨约 600Kb；

2. 基因之间呈一定的保守性, 即同线保守。同线保守广泛存在于与人进化距离较近的小鼠、牛、猪、鸡等动物中。对河豚鱼部分较小区域的基因连锁分析, 显示河豚鱼存在着较高程度的同线保守。一般推测, 由于河豚鱼与人之间的进化距离较远, 连锁打破率(linkage disruption)较高(按照哺乳动物连锁打破率1次/2-3万年计算, 450万年的进化时间里, 河豚鱼约有180次连锁打破, 其平均连锁距离为 $400\text{Mb}/(180+22)=2\text{Mb}$), 所以大片段区域的同线保守效果将会下降, 甚至较小区域的基因顺序与方向也不尽完全一致, 但这并不影响其总的同线保守趋势。

由于河豚鱼基因组具有以上特点, 可作为一新型模式生物发挥以下积极作用:

一、人类基因组计划的参照系:

根据其基因组相对于人来说, 既“全”又“小”, 可与其他模式生物的基因组序列一样作为人类基因组计划中四张图的又一参照系, 这既可避免因大量重复序列的存在而很难搭建“克隆重叠群”的尴尬, 又可省钱、省时、省力。

二、为研究基因的结构、功能和表达、调控找到一简便的坐标:

1. 内含子的研究: 生物界一直对内含子具有怎样的功能的问题至今未有定论。内含子的存在肯定有其积极的意义, 即其具有一定的功能是公认的, 如有些内含子起着增强子的作用或一个基因的内含子可能为另一基因的外显子等等。从内含子中发现基因表达的增强子已成为比较基因学最具有特色的贡献之一。因为增强子是否存在、存在于哪一个内含子中, 无法通过构建不同的表达质粒(即cDNA与任一内含子的不同组合)来验证。通过对不同物种的同源基因的比较, 我们可直观地看到内含子在进化过程中的趋势, 根据“越保守可能越有作用”这一原则来初步找到可能会起作用的“候选内含子”, 进行更深层次的研究, 从而起到事半功倍的效果。

2. 基因重要结构域的发现: 一个基因的所有外显子编码的蛋白质中的所有结构域并不是都起着同样的作用。以前大多数依靠对蛋白质的结构、功能分析, 如抗原抗体反应或通过突变体的功能比较等途径来发现哪些结构域对蛋白质的功能起着决定性的作用。这些方法既费钱、费力, 又费时。根据上述“越保守越重要”的原则, 通过对河豚鱼和人类之间的同源基因结构的电子杂交, 同源性越高的外显子所起的作用可能越大, 同源性越小的作用越小, 如Maheshwar对TSC2基因在人和河豚鱼中进行了比较研究, 发现了功能重要的结构域确实有高度的保守性, 还由此发现了该基因另外三个高度保守的区域, 可能也具有重要的作用, 这当然有待于进一步的研究。但是由于小鼠虽具有普遍高的同源性但基因很大, 而线虫、果蝇虽基因较小但又都具有普遍低的同源性, 这使它们失去了这一重要作用。

3. 基因的表达、调控的研究更趋“自然化”: 人的基因普遍较大, 由于转基

因技术条件的限制,人们通常采用的是利用 cDNA 来研究基因的表达,根本谈不上对基因组实施基因之间相互调控的研究了。这种趋向“简单化”的研究方式必然给实验结果的“原始、真实、可靠”大打折扣。河豚鱼基因的“压缩”与基因之间的“去粗存精”使得在现有技术条件下对单个完整基因甚至是一基因组的转基因工作成为现实,从而实现了基因表达、调控的“自然状态化”,对实验结果产生的影响不言而喻。

上述作用必须基于同源基因的功能不变, Venkatesh B 等人通过转基因小鼠成功的实现了这一设想。

三、在致病基因的克隆研究中发挥重要作用:

目前致病基因的克隆主要采取定位克隆、候选克隆、功能克隆的方法。由于河豚鱼小而紧凑的基因组而使其在致病基因的克隆方面起到积极的补充作用。我们可以设想这样的策略:当致病基因被定位到染色体某一位置后,利用该位置的标记位点作为探针或设计兼并引物来筛查河豚鱼的基因组文库,得到河豚鱼的对应片段。由于河豚鱼基因的高度密集性,即基因平均长度为 7-8kb,这样在人基因组中很大的候选区在河豚鱼的对应区域应很小,通过对河豚鱼片段的序列测定和数据分析,获得该片段中的候选基因,再反过来找到人的同源基因,进而对候选基因进行鉴定,这将会大大加快基因克隆的进程。

人的多药耐药基因的结构、功能和表达、调控

人的多药耐药基因家族有 2 个:MDR1 和 MDR3 (有时又称为 MDR2)。二者均位于 7 号染色体的 7q21.1,以头尾相接的方式连接,基因组跨度为 230Kb,基因间序列约为 34Kb,MDR1 基因约为 120Kb 左右,MDR3 为 74Kb,它们之间的同源性为 85%。

MDR1 基因及其蛋白的结构

MDR1 基因全长约为 120Kb,含有 29 个外显子,其中 27 个编码 P-glycoprotein, 14 个编码 P-glycoprotein 的左侧序列, 13 个编码右侧序列; 28 个内含子,其中 26 个位于基因编码区。基因全长 cDNA 约为 4.5Kb。

MDR1 基因有二个启动子(上游启动子和下游启动子)。上游启动子位于转录起始点上游 70Kb,大多数正常组织的转录是从下游启动子开始的,下游启动子位于外显子 1 中,转录起始位点位于起始密码子上游 140 核苷酸,位于转录起始点 5'端的外显子部分称之为 exon1a,3'端的称之为 exon1b,转录起始密码子 ATG 位于 exon2 中。转录位点有一 CAAT box,但缺少象大多数真核细胞一样的 TATA box;位点下游有两个 GC 富含区可识别转录因子 SP-1,位点上游有一个识别 AP-1

的回文区, 具有增强子的功能; 在 3' 末端终止密码子 TGA 的下游有加尾信号结构。

MDR1 基因编码区的外显子大小范围为 49-587bp (也就是 16-69 个密码子), 其平均大小为 47.5 个密码子, 与其他基因的外显子平均值 44.5 密码子基本相符。所有的剪接位点都遵循 GT/AG 法则。

MDR1 编码的蛋白质 P-glycoprotein 由 1280 个氨基酸组成, N 端未甲基化时分子量为 140Kd, 甲基化后为 170Kd。P-glycoprotein 有 12 个跨膜结构和二对核苷酸结合位点(nucleotide-binding site)。1—637 氨基酸组成了 P-glycoprotein 的 N 端半侧; 638—1280 氨基酸构成了 C 端半侧; N 端和 C 端各有 6 个跨膜结构和一对核苷酸结合位点。

N 端半侧和 C 端半侧对称比较发现 43% 的氨基酸 (263/613) 是一致的, 35% (215/613) 的氨基酸虽然结构不一样, 但功能是一样的; 在 DNA 水平上, 其碱基同源性为 52%。二对核苷酸结合位点, 即 N 端的 529-591 氨基酸和 419-446 氨基酸分别与 C 端的 1174-1236 氨基酸和 1061-1088 氨基酸高度保守一致, 同源性分别为 90.3%(56/62)和 77.8%(21/27)。

MDR1 基因的编码产物 P-glycoprotein 具有下列特征: 1. 依靠 ATP 提供能量才能具有活性; 2. 能降低细胞内的药物浓度; 3. 作用底物多种多样。

通过免疫组化的方法发现在人体的各组织中, MDR1 的表达各不相同。

在各种组织、器官中的表达水平的不同也揭示出 P-glycoprotein 的不同作用;

MDR3 基因及其转录特性

MDR3 基因为 74Kb, 含有 28 个外显子, 其中 27 个有编码功能, 大小范围 55—259bp, 平均为 145bp。第一外显子中含有一富含 GC 的非翻译序列, 有证据表明存在 exon-1。起始密码子位于 exon2 中, 编码的蛋白质与 MDR1 一样含有 12 个跨膜结构和二对核苷酸结合位点, N 端半侧和 C 端半侧互相对称。N 端半侧的 3 和 4 跨膜结构由同一外显子编码, C 端半侧的 9 和 10 跨膜结构亦由同一外显子编码。其余的 8 个跨膜结构由不同的外显子编码; 外显子 12、13、14 和 25、26、27 分别编码 N 端半侧和 C 端半侧的核苷酸结合位点。27 个内含子中含 20 个 0 型、1 个 1 型和 6 个 2 型, 其中只有 9 号和 23 号内含子以及 13 号和 26 号内含子具有相同的位置。

通过对人的 MDR1 和 MDR3 基因的序列比较, 二者从外显子 6 至外显子 27 都具有完全相同的碱基长度和位置。Exon28 在 MDR1 中为 587 个碱基, 在 MDR3 中为 259 个碱基, 这可能是 3' 非翻译区长度不一致造成的。各内含子的大小在 MDR1 和 MDR3 中未见有保守性, 尤其是内含子 4 在 MDR1 中 >20Kb, 而在 MDR3 中为 8Kb。

MDR3 在人体中表达的组织很有限, 仅在肝脏等少数组织中可检测到。至今

我们仍未发现 MDR3 参与多药耐药作用的证据。

由于 MDR1 和 MDR3 基因及其相应的蛋白质结构有着明显的两侧 (3' 和 5', N 端和 C 端) 重复性, 人们设定了一原始基因 (primordial gene) 内复制模式 (internal duplication) 来解释这种结构。1982 年 Higgins、Gilson 等人发现细菌蛋白 hisP、malK、OppD、pstB 都是膜蛋白, 都是依靠 ATP 提供能量而负责将胞浆内的一些物质运输到胞浆外。这些细菌膜蛋白的结构高度相似, 都有 6 个跨膜结构和一对核苷酸结合位点, 其中 hisP 与 MDR1 有 33% 的氨基酸一致, 另有 28% 的氨基酸功能一致, 尤其是核苷酸结合位点的氨基酸在各种细菌蛋白和人的 P-glycoprotein 之间高度同源。鉴于上述特点, 可将这组结构称之为 ABC 超级家族 (ATP-binding cassette superfamily)

通过对各种耐药机制的研究将为临床治疗各种肿瘤、各种传染性疾病找到更好的途径。对于大多数化疗药物来说, 增加其剂量可提高肿瘤的治愈率, 对于一些难治性肿瘤和非霍杰金淋巴瘤, 强烈化疗能获得缓解和提高患者生存率。但由于化疗导致骨髓抑制使许多抗肿瘤药物的使用剂量有限而难以发挥其全部作用, 严重的骨髓抑制可并发感染而不得不在化疗获得效果之前停止给药。已有的一些减轻骨髓抑制的办法如骨髓移植、单克隆抗体导向药物、脂质体包裹药物、区域性化疗、造血因子等等, 都只能作为“姑息疗法”, 并不能从根本上获得大剂量化疗的有效保障, 也就很难获得肿瘤的治愈。鉴于 MDR1 基因的广谱耐药性, 将其转移到造血干细胞后, 可使造血干细胞获得对多种化疗药物的耐受性。将基因转移后的造血干细胞回输给病人, 可进行反复大剂量化疗而不会出现骨髓抑制, 从而提高肿瘤的治愈率和病人的长期生存率。

同时利用 MDR1 基因转移建立的细胞系和动物模型, 将在药学的研究上具有重要的价值, 利用 MDR1 基因的多药耐药特性可使其作为构建某些特殊重组体时的特殊的选择标记。检测肿瘤本身的多药耐药蛋白表达量的高低及在肿瘤治疗中多药耐药表型的改变将有助于预测病情和改变治疗方案。

肿瘤细胞的耐药表型可随着诱导药物的不同、肿瘤细胞的种类、分化阶段的不同而表现出不同的耐药表型, 可以是以某一耐药基因表达, 也可以是多种耐药基因同时表达的多种耐药表型。

本课题的技术路线和结果

本文所采取的策略是: 利用人的 MDR1 全长 cDNA 为探针, 与河豚鱼 cosmid 文库进行异种杂交, 获得可能含河豚鱼多药耐药基因的候选克隆, 对这些候选克隆反复杂交验证后最终获得阳性克隆。限于目前国内的技术力量和经费, 对其中

一阳性克隆先进行亚克隆,即将杂交阳性片段克隆至 pBS 载体中,将获得的阳性亚克隆用 ExoIII 法建立次级亚克隆,对次级亚克隆测序后,再对测序结果进行同源比较,最终获得该 cosmid 含多药耐药基因的确凿证据。然后采用鸟枪法测定该 cosmid 的全长序列;应用生物信息学的一些技术,对全长序列进行分析,主要是分析其剪接结构,获得河豚鱼多药耐药基因的预测 cDNA 序列和完整的阅读框架,再对该 cDNA 序列进行同源性分析和自身结构分析。实验结果表明异种杂交是可行的,河豚鱼中至少含有三个亚型的多药耐药基因,cosmid 124A22 中含有 2 个完整的多药耐药基因。FuguMDR1 的全长序列为 10.8kb,只有人的 1/7-1/12; fuguMDR2 的全长序列为 10.2kb,也只有人的 1/7-1/12,三个基因以头尾相接的方式连接,与人的多药耐药基因组中的基因位置一样,从而具有高度的同线保守性。在每个多药耐药基因中都存在 12 次跨膜结构,重要的结构域即核苷酸结合位点与人或其他生物的多药耐药基因之间呈现高度的保守性。在 cosmid 124A22 中尚存在其他 4 个基因。这样 cosmid 124A22 中的平均基因长度为 6kb,与文献统计数据 7-8kb 吻合。这些结果都有力的支持了河豚鱼作为新型模式生物的理论。

河豚鱼作为模式生物在研究基因的表达、调控方面正在起步,仅仅通过比较基因组学和目前的其他方法尚不足以完全解决基因的调控问题。由于人的基因的全长序列很难获得,更谈不上基因组的分离。本文的序列结果已经显示出我们可以较容易的获得 fuguMDR1 和 fuguMDR2 的全长序列,也可以获得二者的全长基因之和。如果我们分别将这三段序列克隆至表达载体中,通过直接表达后获得二者的全长 cDNA 序列,有了真正的全长 cDNA 后,可以找到真正的阅读框架,从而确定各内含子与人或其他生物对应内含子之间的同源保守,以根据“功能越重要,同源性越高”的原理来寻找可能存在于内含子中的重要功能区,然后通过基因组、全长基因、全长 cDNA 的表达差异来寻找基因的调控结构。另外,通过研究河豚鱼多药耐药基因的耐药谱,与其他生物多药耐药基因的耐药谱相比较,也许可以发现多药耐药基因表达产物与底物结合的一些的有趣线索或发现一些新的药物或毒物的耐受原因。

第一部分:

河豚鱼多药耐药基因的克隆

试剂与材料

材料和试剂

河豚鱼 cosmid 文库 (购自英国人类基因组计划中心)

人的多药耐药基因全长 cDNA 克隆菌株;

人的多药耐药基因全长 cDNA 5'端序列引物 (美国赛百胜公司合成)

primer1 F: CGGGATGGATCTGAAGGG (-4bp—14bp)

primer2 R: AGAACCCCTGTATCATTGAGATC (289bp—310bp)

3'端序列引物

primer3 F: TAGACAGCCTCATATTTTGC (3573bp—3592bp)

primer4 R: AGAGTTCCTGCGCTTTGTTCCAC (3825bp—3848bp)

人基因组 DNA、河豚鱼基因组 DNA、鲑精 DNA

DH5 α 菌种

pBS-SK 菌种

LB 培养基、TB 培养基、SOC 培养基、LB 固体培养基

(以上配方详见《分子克隆》第二版附录 A)

溶液 I 50 mmol/L glucose
 25 mmol/L Tris-Cl (pH 8.0)
 10 mmol/L EDTA (pH8.0)

溶液 II 0.2 mol/L NaOH
 1% SDS

溶液 III 3 mol/L KAc
 2mol/L HAc

变性液 0.5mmol/L NaOH
 1.5mmol/L NaCl

中和液 0.5mmol/L Tris-Cl pH8.0
 1.5mmol/L NaCl

转移液 3mmol/L NaCl

- 0.3mmol/L 柠檬酸钠
- Church 杂交液: 0.5mmol/L 磷酸盐缓冲液 (NaH_2PO_4 、 Na_2HPO_4) pH 7.2;
7%SDS; 1mmol/L EDTA
- 洗脱液 I 2×SSC; 0.1%SDS
- 洗脱液 II 0.1×SSC; 0.1%SDS
- 琼脂糖 (中山公司) 和低熔点琼脂糖 (美国 Bio-Rad 公司)
- 工具酶: XhoI, BclI, HindIII, EcoRI, XbaI, NotI, BamHI, EcoRV, BstXI, SmaI,
T₄ ligase, Klenow 大片段酶, 溶菌酶, CIAP(牛小肠碱性磷酸酶),
ExoIII, S1 核酸酶, Agarose, Rnase, X-gal, IPTG 等
(购自 Promega 公司)
- 抗菌素: 卡那霉素、氨苄青霉素 (华美公司)
- 尼龙膜 (美国 Amersham 公司)
- OLB : 由本室提供
- 同位素 α -³²PdCTP: 亚辉公司
- PCR 纯化试剂盒 (美国 clontech 公司)
- 其他: 乙醇、氯仿、异戊醇、异丙醇、酚、甘油、CsCl、LiCl、盐酸等

主要仪器

- 高速冷冻离心机: 美国 OUTPONT 公司
- 超速离心机: 美国 Beckmen 公司
- 基因导入仪: 美国 Baxter Disgnostics 公司
- 电泳仪、电泳槽 美国 Bio-Rad 公司和北京六一仪器厂
- 恒温水浴: 德国 Jucabo 公司
- 恒温台式离心机: 美国 OUTPONT 公司
- 台式离心机: 美国 Beckmen 公司
- 紫外分光光度计: 美国 Beckmen 公司
- PCR 仪: 美国 PE 公司
- 37°C 培养箱: 湖北黄石医疗仪器厂
- 恒温摇床: 美国 New Brunswicd 公司
- 真空转膜仪: 美国 Pharmacia 公司
- 杂交箱: 美国 Robbins 公司
- 杂交管和杂交网: 美国 Bio-Rad 公司
- ABI377 测序仪

方 法

实验方法及策略:

河豚鱼多药耐药基因的克隆

河豚鱼 cosmid 文库

|
|人 mdrl cDNA 全长序列探针

|
候选克隆

|southern 杂交验证

阳性克隆

|杂交阳性片段的亚克隆

阳性亚克隆

|ExoIII 建立测序亚克隆系统

次级亚克隆

| 测序

| 序列比较、分析

获得河豚鱼 mdrl 阳性 cosmid 克隆及其次级亚克隆和部分序列

一、 探针的制备

1. 用接种环蘸取人的 MDR1-cDNA 菌种, 划线于含 Amp^r (终浓度为 50 μ g/ml) 的 LB 培养皿中培养过夜;
2. 挑取 3 个单克隆分别于 10ml 含 Amp^r 的 LB 培养液中于 37 $^{\circ}$ C 摇床中剧烈震荡过夜;
3. 质粒提取:
 - a. 4 $^{\circ}$ C 以 6000 转/分将 10ml 菌液离心 10min 以沉淀细菌;
 - b. 弃上清, 在沉淀中加入 200 μ l 溶液 I, 反复吸打, 以重悬细菌沉淀, 将悬浮液转移到 1.5ml 离心管中;
 - c. 加入新鲜配置的溶液 II, 立即颠倒离心管数次, 以充分混匀, 置冰上 3min (切勿震荡);
 - d. 加入 300 μ l 冰预冷的溶液 III, 立即颠倒离心管数次, 充分混匀溶液, 形成棉絮状沉淀, 置冰上 10min;
 - e. 用微量离心机于 4 $^{\circ}$ C 以 12000rpm 离心 5min, 将上清转移到另一离心管中;
 - f. 加入 Rnase (终浓度为 50 μ g/ml), 将离心管置于 37 $^{\circ}$ C 水浴中 30min ;

- g. 加入等体积的酚: 氯仿: 异戊醇, 充分混匀, 用微量离心机于 4°C 以 12000rpm 离心 5min。转移上清到另一离心管中;
- h. 在转移的上清中加入等体积的异丙醇, 室温放置 10min;
- i. 用微量离心机于 4°C 以 12000rpm 离心 15min, 沉淀 DNA;
- j. 弃上清, 将离心管倒置于吸水纸上, 使溶液全部流尽;
- k. 在 DNA 沉淀中加入 70%乙醇, 弹散沉淀, 于 4°C 以 12000rpm 离心 10min, 重新沉淀 DNA ;
- l. 弃上清, 将离心管倒置于吸水纸上, 在室温自然干燥 DNA 沉淀;
- m 在沉淀中加入 20 μ l TE 溶解 DNA, 取 1 μ l 于 0.8%的琼脂糖凝胶中电泳鉴定, 其余的置于-20°C 保存。

4. 质粒验证

将提取的三个质粒以 XhoI 酶切, 酶切体系如下:

DNA	1 μ l
XhoI	0.5 μ l (5U)
10 \times buffer	2 μ l
dH ₂ O	16.5 μ l

TV(total volume) 20 μ l

置于 37°C 水浴 2hr。取 5 μ l 于 0.8% 琼脂糖凝胶中电泳鉴定。

5. 插入片段的分离

取验证正确的人 MDR1-cDNA 质粒 30 μ g, 先以 SacII 酶切, 酶切体系

DNA	20 μ l
SacII	15 μ l (150U)
10 \times buffer	20 μ l
dH ₂ O	143 μ l

TV 200 μ l

置于 37°C 水浴 2hr, 取 2 μ l 于 0.8%琼脂糖凝胶鉴定, 酶切完全后, 在酶切体系中加入 200 μ l 酚: 氯仿: 异戊醇, 混匀后于 4°C 以 12000rpm 离心 10min, 将上清转移到另一离心管中; 加入 3 倍体积的无水乙醇, 充分混匀, 置于室温 5min, 于 4°C 以 12000rpm 离心 15min, 沉淀 DNA; 弃上清, 在沉淀中加入 1ml 70%乙醇洗涤 DNA, 于 4°C 以 12000rpm 离心 15min; 弃乙醇, 室温干燥沉淀, 将沉淀溶于 143 μ l 去离子水中, 加入 15 μ l(150U)XhoI, 20 μ l 10 \times buffer, BSA 2 μ l, 充分混匀, 置于 37°C 水浴 2hr, 取 2 μ l 电泳鉴定酶切完全后, 准备回收插入片段。

6. 插入片段的回收

- a. 将上述酶切产物于 0.8% 的低熔点凝胶在 TAE 电泳液中电泳分离;
- b. 将 4.3kb 的片段从凝胶中挖出, 置于 1.5ml 离心管中;
- c. 称量凝胶质量 (1g=1ml), 加入 0.04 体积的 25× Agarase 缓冲液, 将离心管置于 65°C 水浴中 15min, 直到凝胶全部融化;
- d. 待凝胶液冷却到 45°C 时加入 5U agarase (1U/100μl 凝胶液), 充分混匀, 置于 45°C 水浴中 1hr;
- e. 加入 0.1 体积的 3M NaAc (pH5.5), 混匀, 置于冰上 15min;
- f. 于 4°C 以 12000rpm 离心 15min, 沉淀寡糖;
- g. 取上清液, 加入 3 倍体积的无水乙醇置于室温 5min, 于 4°C 以 12000rpm 离心 15min 以沉淀 DNA 片段;
- h. 以 1ml 70%乙醇洗涤沉淀, 离心, 室温干燥沉淀, 加入 20μl 去离子水溶解沉淀;
- i. 取 1μl 电泳鉴定, 同时取 1μl 测定其 OD₂₆₀ 值以确定 DNA 片段的浓度。

二、 预实验

1. 杂交膜的制备

- a. 分别取 30μg 人基因组 DNA 和 15μg 河豚鱼基因组 DNA, 稀释于 420μl 去离子水中, 加入 50μl 10×buffer, 于 4°C 放置 5hr;
- b. 使一 Tip 温和的搅拌 DNA 溶液;
- c. 加入 15μl (150U) HindIII 于 4°C 温和的搅拌 5min, 置于冰上 1hr 后, 再置于 37°C 水浴 1hr;
- d. 再加入 15μl HindIII, 重复上述步骤, 置于 37°C 水浴酶切过夜;
- e. 取 5μl 于 0.8%凝胶中, 电泳鉴定酶切程度;
- f. 在酶切完全的基因组 DNA 中加入 3 倍体积无水乙醇, 置于室温 2min 后, 于 4°C 以在 12000rpm 离心 15min, 弃上清, 以 1ml 70%乙醇洗涤沉淀, 再次离心, 室温干燥沉淀, 加 20μl TE 溶解沉淀,
- g. 将酶切产物于 0.8%凝胶中 4°C 电泳过夜 (电压 1V/cm);
- h. 将凝胶照相定位后, 在真空转膜仪上放置凝胶大小的尼龙膜 (以 2×SSC 湿润), 在尼龙膜上放置凝胶, 打开真空泵, 胶面滴加 0.25N HCl 脱嘌呤 6min, 弃去 HCl; 滴加变性液以变性 15min; 弃变性液, 滴加中和液中和 15min; 弃中和液, 滴加 20×SSC, 转移 1hr 后, 将凝胶移出以 EB 复染, 同时以 2×SSC 冲洗尼龙膜; 将洗好的尼龙膜以滤纸包好, 置于 80°C 烤箱烘干 2hr。

2. 预杂交

将尼龙膜以 2×SSC 湿润后, 放入杂交管中, 加入 10ml church 杂交液, 放入杂交箱中, 将 1mg 超声打碎的鲑精 DNA (片段大小为 500bp 左右) 煮沸 5min,

立即置于冰上 2min, 加入预杂交液中, 以 60°C 预杂交 5hr;

3. 杂交

探针标记: 在 35 μ l 去离子水中加入 2 μ l cDNA 探针 (40ng/ μ l), 煮沸 10min, 置于冰上 5min, 加入 2 μ l α -³²PdCTP、10 μ l OLB、1 μ l Klenow 酶, 混合均匀 (冰上操作), 置于 37°C 水浴 2hr, 检测标记效率。

杂交: 将标记好的探针煮沸 10min, 置冰上 5min, 加入预热至 60°C 的 10ml church 杂交液中, 混合均匀后, 加入到杂交管中, 放入杂交箱中 60°C 杂交过夜。

4. 洗膜、压片

将杂交膜取出, 放入洗脱液 I 中室温漂洗 15min 后, 重新放入杂交管中加入洗脱液 II, 置于 65°C 杂交箱中漂洗 30min 左右, 其间反复检测杂交膜上的同位素信号强度, 至合适的程度后, 用滤纸吸干尼龙膜, 包以保鲜膜, 置于增感屏中于 -70°C 冰箱中压片 24hr, (不同的信号强度决定了压片的时间长短) 冲洗 X 光片。

5. 杂交膜的处理

将杂交膜放入 100°C 的 0.1 \times SSC 中, 自然冷却至室温, 检测膜上的信号是否完全消除, 将膜用滤纸包好, 置于室温保存。

6. 杂交条件的选择

将预杂交温度改为 50°C、55°C 和 65°C, 重复上述预杂交和杂交过程, 选择最佳杂交条件。

三、河豚鱼 comid 文库的杂交

1. 预杂交

将河豚鱼 comid 文库的五张 22 \times 22cm 杂交膜以 2 \times SSC 湿润, 在玻璃板上将杂交膜叠放, 彼此间隔以杂交网, 抹平, 尽量不留气泡, 将卷好的杂交膜-杂交网放入杂交管中, 加入 20ml church 杂交液 (含 100 μ g/ml 鲑精 DNA) 于 60°C 预杂交 5hr。

2. 杂交

探针标记: 标记程序同前, 标记体系如下:

cDNA	8 μ l (40ng/ μ l)
dH ₂ O	140 μ l
α - ³² PdCTP	8 μ l (80 μ ci)
OLB	40 μ l
Klenow 酶	4 μ l

TV 200 μ l

杂交过程同前。

3. 洗膜、压片

先以洗脱液 I 室温漂洗杂交膜 15min, 将杂交膜卷好后重新放入杂交管中, 加入洗脱液 II,

于 65°C 漂洗 30min-60min, 其间反复检测杂交膜上的杂交信号强度, 至合适强度后, 吸干, 包以保鲜膜。-70°C 压片 24hr, 冲洗 X 光片。

4. 杂交膜的再处理同前。

四、 杂交阳性信号克隆的挑选和鉴定

1. 根据杂交膜的点膜规则, 筛选两点同为阳性信号的克隆, 按其信号强弱顺序购买信号较强的 34 个克隆 (根据此套杂交膜的说明, 除外了假阳性克隆后);
2. 蘸取每个 cosmid 克隆, 划线于含卡那霉素 30 μ g/ml 的 LB 固体培养基上;
3. 从每个培养板上各挑取 3 个单菌落, 分别加入到 5ml LB 培养液中, 于 37°C 剧烈震荡过夜,
4. 取 1ml 培养液于 1ml 30%高压过的甘油中, 混匀, 置于-70°C 冰箱中保种, 其余的使用碱裂解法提取 cosmid, 电泳鉴定;
5. 以 EcoRI 酶切鉴定各 cosmid, 电泳鉴定。

五、 候选克隆的再次筛选

1. 斑点杂交膜的制备

将一大大小合适的尼龙膜以 2 \times SSC 湿润后, 置于真空狭缝转移装置中, 固定好该装置, 将经上一步鉴定准确的各 cosmid (约 500ng) 以 EcoRI 酶切后加入狭缝中, 启动转移装置, 直至所有的 cosmid 溶液完全被吸干至尼龙膜上, 将转有 cosmid DNA 的尼龙膜置于含变性液的平皿中 15min, 含中和液的平皿中 15min, 滤纸吸干, 置于 80°C 烘箱中烘干 2hr 以固定 DNA。

2. Southern 杂交膜的制备

将各 cosmid DNA (约 1 μ g) 以 EcoRI 酶切后, 于 0.7%的凝胶 4°C 电泳过夜; 将一大大小合适的尼龙膜以 2 \times SSC 湿润后, 置于真空转膜仪上, 覆以凝胶, 打开真空泵, 胶面滴加变性液抽吸 15min; 去除变性液, 滴加中和液, 抽吸 15min; 去除中和液, 滴加 20 \times SSC, 抽吸 1hr, 关闭装置, 将胶取出置于 EB 溶液中复染; 将尼龙膜标记后取出, 以 2 \times SSC 漂洗数次, 滤纸吸干, 置于 80°C 烘箱中烘干 2hr 后备用。

3. 杂交验证

预杂交: 上述二张尼龙膜以 2 \times SSC 湿润后, 彼此隔以杂交网, 卷好后置于杂交管中, 将以变性的鲑精 DNA 1mg 与 10ml church 杂交液混匀后加入杂交管中, 于 60°C 预杂交 5hr;

探针标记程序及杂交过程同前, 标记反应体系如下:

cDNA	80 ng (2 μ l)
dH ₂ O	35 μ l
OLB	10 μ l
α - ³² PdCTP	2 μ l
Klenow 酶	1 μ l

TV 50 μ l

洗膜、压片、膜处理过程同前。

六、 全长 cDNA 5'端和 3'端探针的制备

设计、合成 5'端和 3'端各一对引物，以 MDR1 全长 cDNA 为模板，进行 PCR，扩增获取 5'端和 3'端一段序列作为探针，PCR 反应体系如下：

cDNA	1 μ l (5ng)	$\times 5$
10 \times buffer	2.5 μ l	$\times 5$
dNTP mix	1 μ l	$\times 5$
primer1(3)	1 μ l (25pmol)	$\times 5$
primer2(4)	1 μ l (25pmol)	$\times 5$
Taq 酶	1 μ l (2U)	$\times 5$
dH ₂ O	17.5 μ l	$\times 5$

TV 25 μ l $\times 5$

反应程序: 95 $^{\circ}$ C 5min

95 $^{\circ}$ C 30s; 58 $^{\circ}$ C 30s; 72 $^{\circ}$ C 30s 35 个循环

72 $^{\circ}$ C 10min

扩增产物在 TAE 缓冲液中电泳后，将片段从凝胶中切下，称量其质量 (1g=1ml) 加入 3 倍体积的凝胶溶解液，于 55 $^{\circ}$ C 加热 10min，直至凝胶完全溶解，加入 10 μ l 玻璃奶，混匀，置于室温 5min，其间不时摇动以保证玻璃奶与 DNA 充分结合，于 12000rpm 室温离心 5s，去除上清，加入 1ml 洗涤液，吸打溶液，使玻璃奶散开，12000rpm 室温离心 5s，去除上清，室温干燥玻璃奶，加入 20 μ l 去离子水于玻璃奶中，混匀，室温搁置 5min，以 12000rpm 离心 1min，吸出上清 DNA 溶液；

测定 PCR 回收产物的 OD 值，备后用。

七、 含有完整 MDR 基因的阳性克隆及其酶切阳性片段的筛选与鉴定

将上一步获得的 9 个阳性克隆分别以 EcoRI 和 HindIII 酶切后，电泳、转膜；分别以全长 cDNA、5'端序列和 3'序列为探针，进行 Southern 杂交。

八、 阳性片段的亚克隆

1. 电穿孔感受态细胞的制备

用接种环蘸取 DH5 α 菌种，涂于 LB 培养基（不含抗生素），置于 37 $^{\circ}$ C 温箱中培育 16hr，挑取单克隆至 5ml LB 培养基中，37 $^{\circ}$ C 剧烈震荡过夜；

将 500 μ l 上述菌液加入 500ml LB 培养液中，37 $^{\circ}$ C 剧烈震荡，直至 OD₆₀₀=0.5-0.6；将 500ml 菌液置于冰上 10min，4000rpm/4 $^{\circ}$ C/10min/GS3 离心，弃上清，将菌体重悬于冰预冷的 500ml 高压灭菌后的 10%甘油溶液中，置于冰上 5min；5000rpm/4 $^{\circ}$ C/10min GS3 转头离心，弃上清；重复上一步；用 40ml 冰预冷的 10%甘油重悬细胞；6000rpm/4 $^{\circ}$ C/SS34 离心，弃上清；用少量 10%甘油重悬细胞，使终体积为 2ml；200 μ l 一份分装，置于 -70 $^{\circ}$ C 冰箱中保存。

2. 插入片段的制备

将上述三个探针杂交后均含有阳性片段的克隆之一 124A22 以 HindIII 酶切后，于 0.7%的 TAE 低熔点凝胶中 4 $^{\circ}$ C 过夜，将其中的杂交阳性片段 1、2、3、6、7、10 从凝胶中挖出，以 PCR 纯化 kit 回收；

3. 载体的制备

挑取 pBS-SK 菌种于 10ml LB 培养液中，剧烈震荡过夜，以 PEG 法提取 pBS 质粒（方法详见测序质粒提取）。

将 1 μ g pBS 质粒以 Hind III 酶切完全后，以 PCR 纯化 kit 回收，溶于 10 μ l 去离子水中，按下列体系进行去磷酸化处理

DNA	10 μ l
10 \times buffer	2 μ l
CIAP	1 μ l(1U)
DH ₂ O	7 μ l

TV 20 μ l

于 37 $^{\circ}$ C 水浴中温育 30min，温育结束后，加入 EDTA (pH8.0) 至终浓度 5mmol/L，65 $^{\circ}$ C 加热 1hr 以灭活 CIAP，用 PCR 纯化 kit 回收 DNA，溶于 20 μ l 去离子水中。

检测 pBS 酶切完全程度和去磷酸化效率，建立下列并列连接体系：

DNA	酶切后 pBS	酶切后 pBS	去磷酸化 pBS	去磷酸化 pBS	未酶切 pBS
T ₄ ligase	+	-	+	-	-
10 \times buffer	+	+	+	+	+
dH ₂ O	x	x	x	x	x

16 $^{\circ}$ C 连接过夜。

取 1-2 μ l 连接产物, 与 20 μ l 电穿孔感受态细胞混匀后加入电穿孔杯中进行转化, 电穿孔的条件为: 电极间距为 0.2cm, 电容为 25 μ F, 电阻为 200 Ω , 电压为 2500V, 每次电穿孔完毕后, 立即向电穿孔杯中加入 980 μ l SOC 培养基, 吸出后于 37 $^{\circ}$ C 缓慢震荡复苏, 于 4 $^{\circ}$ C 以 5000rpm 离心 1min, 去除 800 μ l 上清, 留下 200 μ l 重悬细胞, 涂于含 50 μ g/ml 氨苄青霉素、表面涂有 40 μ l x-gal (20mg/ml) 和 4 μ l IPTG (200mg/ml) 的 LB 培养基上, 于 37 $^{\circ}$ C 培养 16hr。

4. 阳性片段亚克隆

将回收的阳性片段和处理好的 pBS 载体按摩尔比为 2 的连接体系于 16 $^{\circ}$ C 连接过夜; 取 1-2 μ l 连接产物, 以电穿孔法转化 (方法如前), 铺板; 挑取单克隆, 少量培养后, 保种, 提取各亚克隆质粒, 酶切鉴定。

九、各亚克隆的大量提取、纯化

1. 将各鉴定好的亚克隆菌种分别加入到对应的 500ml TB 培养液中, 于 37 $^{\circ}$ C 剧烈震荡过夜;
2. 用 GS3 转头于 4 $^{\circ}$ C 以 5000rpm 离心 15min, 弃上清, 以 100ml 冰预冷的 STE 重悬细胞沉淀, 再次离心, 弃上清;
3. 将洗过的细菌沉淀重悬于 10ml 溶液 I 中;
4. 加入 1ml 新鲜配制的溶菌酶溶液;
5. 加入 20ml 新配制的溶液 II, 缓慢颠倒离心管数次, 以充分混匀, 于室温放置 3min;
6. 加入 15ml 溶液 II, 颠倒离心管数次, 置于冰上 10min;
7. 用 GS3 转头于 4 $^{\circ}$ C 以 5000rpm 离心 15min, 不开刹车使转头自然停转;
8. 将上清转移至另一 100ml 离心管中, 加入 0.6 倍体积的异丙醇, 充分混匀, 于室温放置 10min, 室温以 6000rpm 离心 15min, 回收质粒;
9. 弃上清, 加入 20ml 70%乙醇洗涤沉淀, 离心、去除乙醇, 加入 3ml TE 溶解沉淀;
10. 将 DNA 溶液加入到 3ml 用冰预冷的 5mol/LiCl 溶液中, 充分混匀, 于 4 $^{\circ}$ C 以 12000rpm 离心 10min;
11. 将上清转移到另一离心管中, 加入等量的异丙醇, 充分混匀, 室温放置 10min, 于室温以 12000rpm 离心 15min 回收沉淀, 再以 70%乙醇洗涤沉淀, 将沉淀溶于 8ml TE 中;
12. 在 8ml DNA 溶液中加入 8.8g CsCl 和 0.8ml 10mg/ml 的 EB, 溶解混匀后, 转移至超离管中, 用石蜡油填充超离管剩余的部分, 封管, Ti50 转头于 20 $^{\circ}$ C 以 45000rpm 离心 48hr;
13. 将一注射用针头插入超离管顶端, 以另一 10ml 注射器插入到条带下约 1cm 处, 回收条带;

14. 在回收 DNA 溶液中, 加入等体积的异戊醇, 震荡混匀后, 室温 4000rpm 离心 5min, 将下层水相转移到另一离心管中, 重复上述步骤 4-6 次, 直至粉红色从水相和有机相中消失;
15. 用 3 倍体积的去离子水稀释 DNA 溶液, 于 4°C 以 3 倍总体积的无水乙醇沉淀 DNA 15min, 再于 4°C 以 12000rpm 离心 15min, 弃上清, 70% 的乙醇洗涤沉淀, 最后将沉淀溶于 100 μ l 去离子水中;
16. 测定各 DNA 的 OD₂₆₀ 值, 计算其浓度, 置于 -20°C 中保存备用。

十、 各亚克隆的 ExoIII 次级亚克隆的建立

1. 限制性内切酶的选择

将 pBS 插入区的各种限制性内切酶酶切各亚克隆, 选择只有一个酶切位点的限制性内切酶进行下列实验。

2. 以 BamHI 酶切含 1 号插入片段的亚克隆 140 μ g, 酶切完全后以酚: 氯仿: 异戊醇抽提, 加入 3 倍体积的无水乙醇沉淀质粒 DNA, 70% 乙醇洗涤沉淀后, 以 50 μ l Klenow 1 \times buffer 溶解沉淀;
3. 加入 4 种 α -³⁵SdNTP 的混合物(终浓度各为 40 μ M)和 Klenow 聚合酶(终浓度为 50U/ml), 置 37°C 水浴 10min;
4. 70°C 水浴 10min 以灭活 Klenow 酶, 以酚: 氯仿: 异戊醇抽提后, 加入 3 倍体积的无水乙醇沉淀, 再次回收质粒 DNA;
5. 将上述质粒 DNA 再以 EcoRV 酶切, 待酶切完全后, 以酚: 氯仿: 异戊醇抽提, 加入 3 倍体积的无水乙醇沉淀、回收质粒 DNA;
6. S1 酶混合物的配制: 170 μ l dH₂O 中加入 27 μ l 7.4 \times S1 缓冲液和 3 μ l S1 酶 (60U), 混匀, 分别取 7.5 μ l 混合物于 24 个 0.5ml Eppendorf 管中, 置于冰上备用;
7. 将上一步回收的双酶切质粒 17 μ g, 加入 6 μ l 10 \times ExoIII 缓冲液和适量的去离子水, 使终体积达到 60 μ l, 置于 37°C 水浴, 5min 后取出 2.5 μ l 加入第一管 S1 酶混合物中;
8. 加入 3 μ l ExoIII(600U)于反应管中, 迅速混匀, 置于 37°C 水浴, 每隔 1min 取 2.5 μ l 于相应的 S1 酶混合物中混匀, 直至全部反应液被取尽;
9. 将所有的 S1 酶反应管从冰浴中取出, 置室温 30min。每管中加入 1 μ l S1 酶终止液 (含 0.3mmol/L Tris-Cl pH8.0, 0.05mmol/L EDTA), 于 70°C 水浴 10min 以灭活 S1 酶;
- 10 回收纯化每管中的 DNA(玻璃奶法). 溶于 10 μ l dH₂O 中, 取 5 μ l 电泳鉴定。在剩下的 5 μ l DNA 溶液中加入 5 μ l Klenow 酶混合物(含 1 μ l 10 \times Klenow 酶缓冲液, 0.5U Klenow 酶), 置 37°C 水浴 3min 后取出, 置于冰上, 加入 1 μ l dNTP(终浓度各为 12.5 μ M), 置 37°C 水浴 5min 后于 65°C 水浴 10min 以灭

活 Klenow 酶;

- 11 每管 DNA 溶液中加入 4 μ l T₄ 连接酶混合物(含 1.5 μ l 10 \times 连接酶缓冲液, 1U T₄ 连接酶) 于 16 $^{\circ}$ C 连接过夜;
- 12 取 1 μ l 连接产物于 20 μ l 感受态细胞中, 以电穿孔法转化, 铺板;
- 13 从代表每一时间间隔的各个平板上各挑取 8 个单克隆, 加入 96 孔板中(每孔加有 200 μ l 含氨苄青霉素 50 μ g/ml 的 TB 培养基)。将 96 孔板置于 37 $^{\circ}$ C 培养箱中培养 48hr;
- 14 从每孔中取 100 μ l 菌液于 0.5ml Eppendorf 管中(96 孔板中每孔再加入 100 μ l 30%甘油, 置-70 $^{\circ}$ C 冰箱中保种), 离心, 弃上清;
- 15 加入 50 μ l 10mM EDTA(PH 8.0)悬浮沉淀, 加入 500 μ l 新配制的 2 \times 裂解液(每 50ml 裂解液含 5M NaOH 2ml, 10% SDS 2.5ml, 葡萄糖 10g), 混匀, 置 70 $^{\circ}$ C 水浴 10min, 冷却至室温。加入 1.5 μ l 4M KCl 和 1 μ l 0.25%溴芬蓝, 混匀, 置冰浴 5min, 以 12000rpm 4 $^{\circ}$ C 离心 3min。每管取 40 μ l 上清电泳、鉴定;
- 16 用同样的步骤处理 2、3、6、7 亚克隆。

十一、 序列测定及分析

1. 挑取各鉴定好的次级亚克隆于 5ml TB 培养液中 (含氨苄青霉素) 于 37 $^{\circ}$ C 剧烈震荡过夜, 取 1.5ml 菌液置于离心管中离心 1min, 弃上清, 加入 200 μ l 溶液 I 悬浮细菌沉淀, 加入 300 μ l 溶液 II, 迅速混匀数次, 置于冰上 5min, 加入 300 μ l 溶液 III, 摇匀数次, 置于冰上 10min, 12000rpm/4 $^{\circ}$ C 离心 10min 后, 转移上到另一离心管中, 加入 RnaseA 至终浓度 20 μ g/ml, 于 37 $^{\circ}$ C 水浴 30min, 加入 800 μ l 氯仿混匀, 离心 1min, 转移上清 (该氯仿抽提过程重复一次), 再加入等体积的异丙醇, 离心 10min, 弃乙醇, 室温干燥沉淀物, 在沉淀中依次加入 32 μ l 去离子水、8 μ l 4M NaCl 和 40 μ l 13%PEG8000, 混匀, 置于冰浴 20min 后于 4 $^{\circ}$ C 以 12000rpm 离心 30min。弃上清, 70%乙醇洗涤沉淀, 室温干燥沉淀, 加入 10 μ l 去离子水溶解沉淀, 置于-20 $^{\circ}$ C 保存;
(以上离心除注明外, 均为室温 12000rpm)
2. 将各次级亚克隆以载体引物 T₃ 或 M13 测序, 测序仪为 ABI377;
3. 将测序结果送至数据库中同源比较、分析。

结 果

- 一、 人MDR1 cDNA 质粒插入片段的获取：经过 SacII 和 XhoI 两次酶切后，可得到的插入片段，即人的MDR1 的全长 cDNA 约为 4.3kb。
- 二、 预实验的杂交结果显示：人类基因组和河豚鱼基因组都含有若干条杂交阳性条带，这说明河豚鱼基因组中含有与人MDR1 基因同源的基因。
预实验显示：当杂交温度为 65°C 时，杂交信号明显减弱，有多条带不很清楚，甚至消失，但杂交温度为 50°C 时杂交背景较高，所以本实验选择杂交温度 60°C 作为人—河豚鱼异种杂交温度。
- 三、 河豚鱼 cosmid 文库杂交后共得到 44 个阳性信号（图 1、2、3、4）（去除膜制作者提供的假阳性记录后）。这些阳性信号强弱相差较大，故选择购买其中的 34 个克隆（带*者为购买的克隆）。

F003I09	F004E03*	F007E01	F007M11*	F012C08*	F020M22*
F023J04*	F027H12*	F029I20*	F030B05*	F036N23*	F037G16
F038M19*	F071H20*	F072G13*	F074F11*	F074L23*	F076L20*
F077E22*	F077F21*	F077O17*	F096G17*	F012I20	F124A21*
F124A22*	F124D21*	F127I02	F130A03*	F133J12*	F137A16*
F137E16*	F137P10*	F141H05	F142C24*	F142N21*	F147N19
F148F17*	F168O13	F186P10*	F190E17*	F199I03*	F199P17*

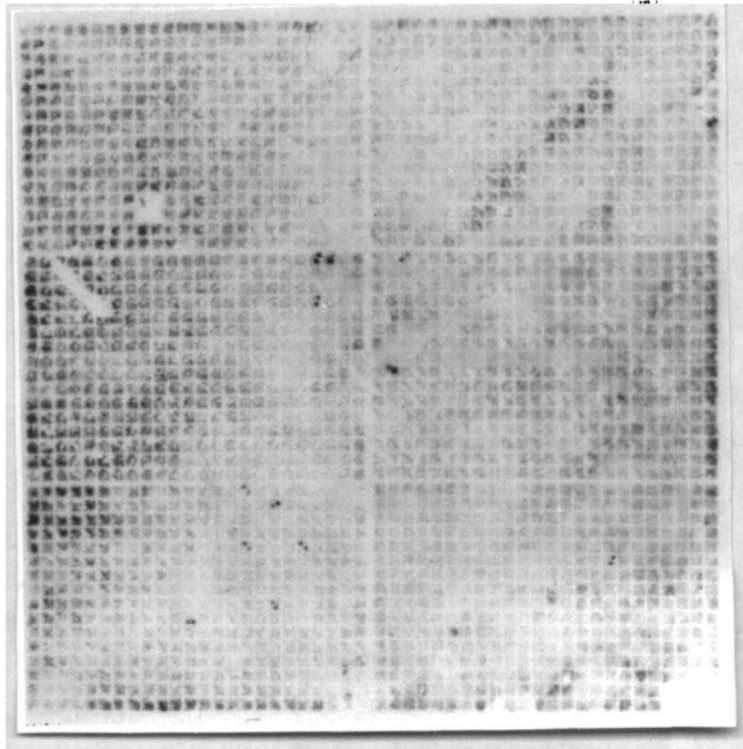


图 1：河豚鱼 cosmid 文库杂交图



图2: 阳性信号 124A21 (→)、124A22 (➤)

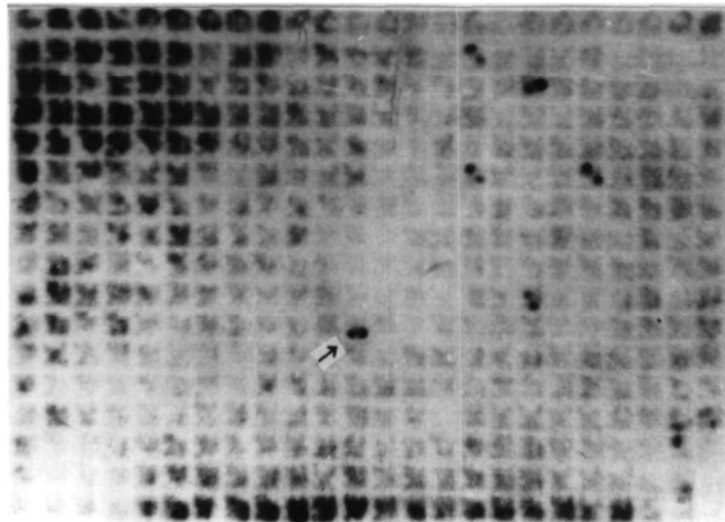


图3: 阳性信号 133J12 (→)

- 四、 各购买的克隆经过酶切验证后, 即每个克隆的至少三个菌落的酶切图完全一致时, 才确认该克隆是所需的候选克隆。
- 五、 候选克隆的斑点杂交和 Southern 杂交验证(图 5、6)表明: 12C8、124A21、124A22、133J12、142N21、148F17、186P10、190E17、199I3、199P17 为阳性克隆, 其中 124A21 和 124A22、133J12 的酶切图(图 7)和杂交图完全一致, 124A21 和 124A22 可能为相邻孔之间的污染所致, 故选择其中之一 124A22 继续进行验证。



图4: 阳性克隆 199I3 (→)、199P17 (>)

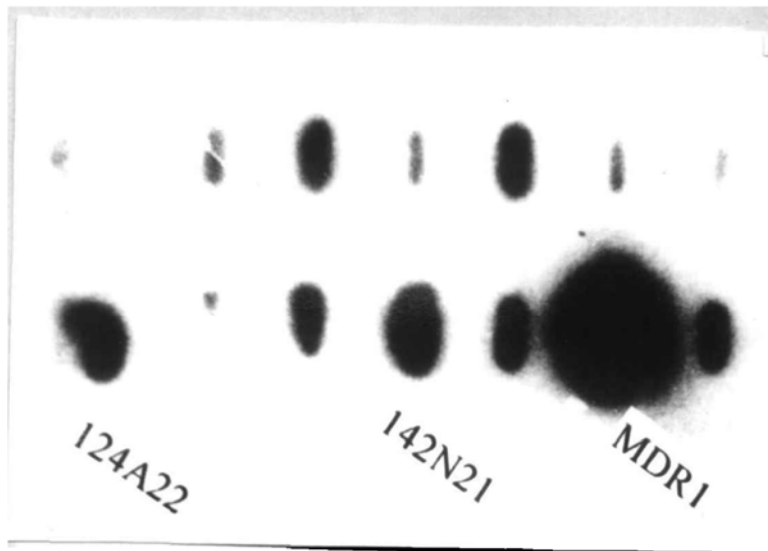


图5: 候选克隆的斑点杂交图

六、 PCR 结果为: 5'端序列 314bp; 3'端序列为 275bp。

七、 将选定的 9 个阳性克隆分别以 EcoRI 和 HindIII 酶切后, 电泳、转膜、杂交, 以全长 cDNA 为探针杂交后 (图 8、9) 证实: 124A22 和 133J12 实际上为同样的克隆; 一般以 EcoRI 酶切后的杂交阳性片段较大而以 HindIII 酶切后的阳性片段较小, 故选择 HindIII 酶切后的阳性片段作为下一步亚克隆的插入片段。杂交结果同时显示 9 个克隆的酶切图和杂交图彼此之间有许多相同大小的片段, 这表明这些克隆中都含有河豚鱼基因组中同样的一段序列的一部分, 只是该序列在不同克隆中长度有差别。

以 5'端序列为探针杂交 (图 10), 186P10 有明显的阳性条带, 124A22、

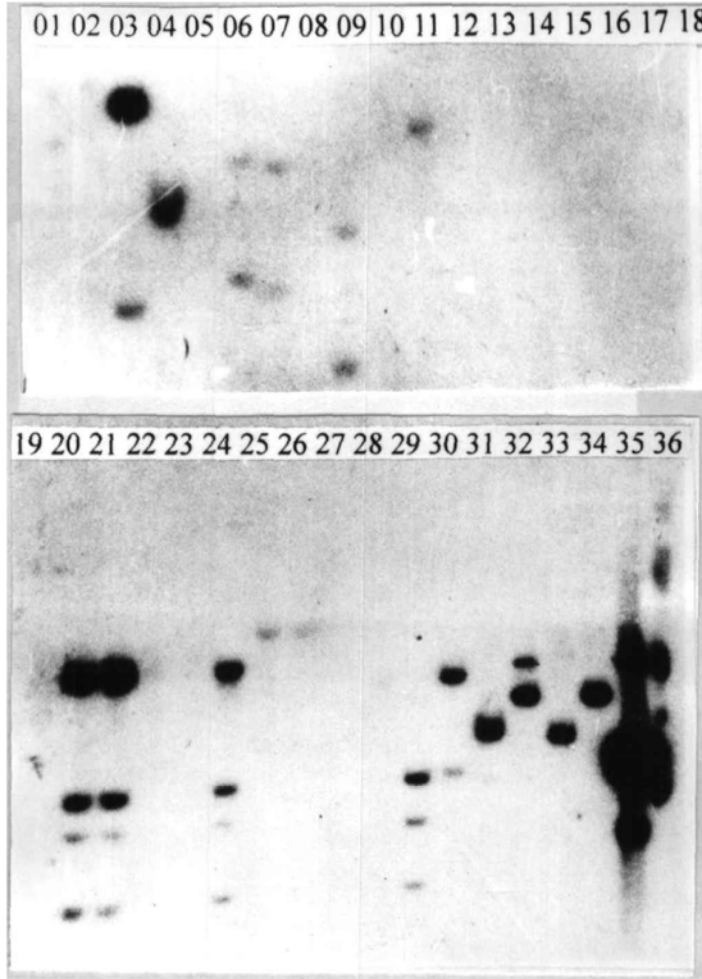


图 6: 候选克隆的 Southern 杂交图

03: 012C08; 20:124A21; 21:124A22; 24:133J12; 29:142N21;
30:148F17; 31:186P10; 32:190E17; 33:199I03; 34:199P17

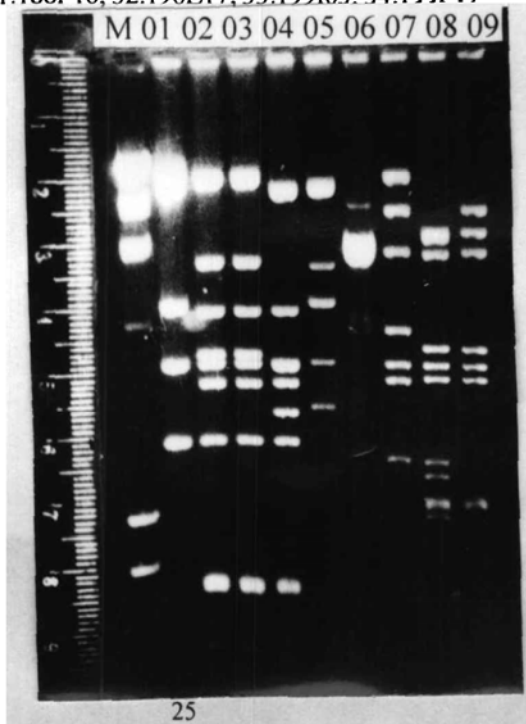


图 7: 阳性克隆的 EcoRI 的酶切图谱

M:λDNA/HindIII; 01:012C08; 02:124A22; 03:133J12; 04:142N21;
05:148F17; 06:186P10; 07:190E17; 08:199I03; 09:199P17

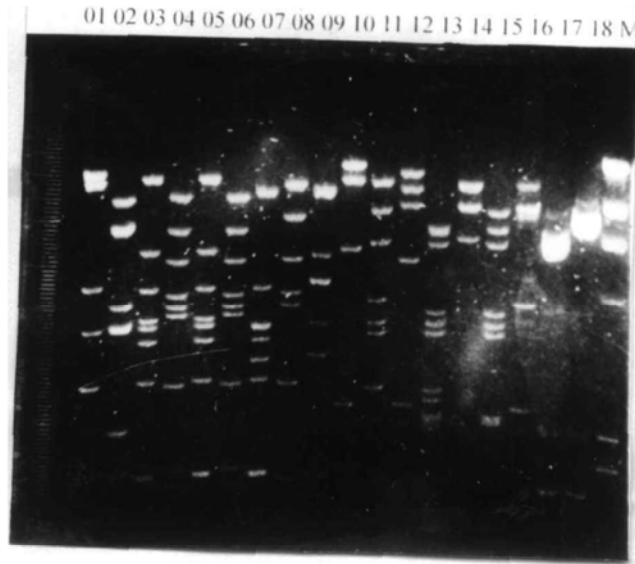


图 8: 阳性克隆的 EcoRI 和 HindIII 的图谱

EcoRI 酶切: 01:012C08; 03:124A22; 05:133J12; 07:142N21;
09:148F17; 11:190E17; 13:199I03; 15:199P17; 17:186P10
HindIII 酶切: 02:012C08; 04:124A22; 06:133J12; 08:142N21;
10:148F17; 12:190E17; 14:199I03; 16:199P17; 18:186P10
M: λDNA/HindIII

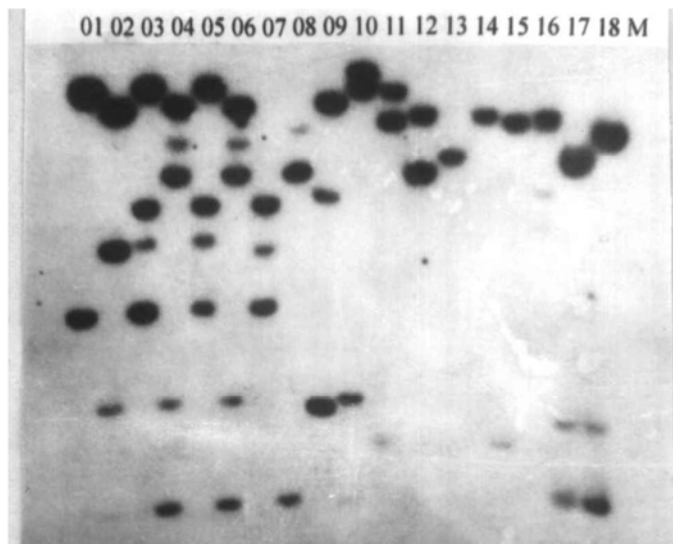


图 9: 以全长 cDNA 为探针的 southern 杂交结果 (序号同图 8)

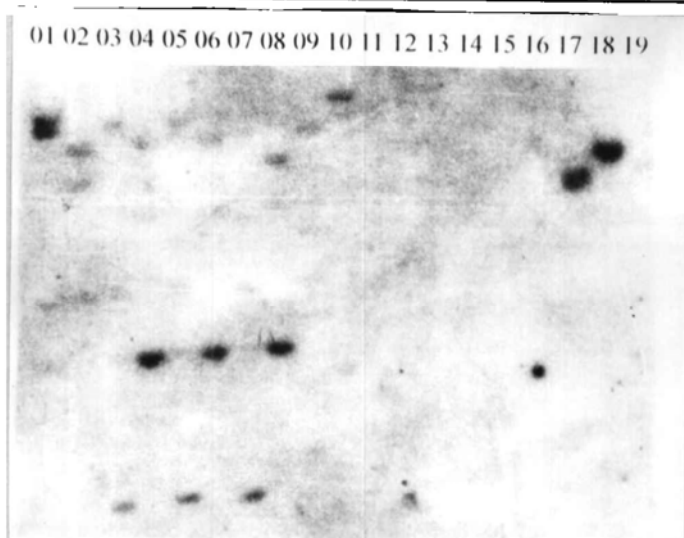


图 10: 以 5'端序列为探针的 southern 杂交结果 (序号同图 8)

133J12、142N21 有一条较小的杂交阳性片段。

以 3'端序列为探针杂交 (图 11) 后, 大部分杂交信号与全长 cDNA 杂交信号一致, 但信号强度较弱。

三次杂交结果综合表明: 124A22 和 133J12 都可能含有河豚鱼的全长多药耐药基因, 故选择 124A22 进行下一步实验。

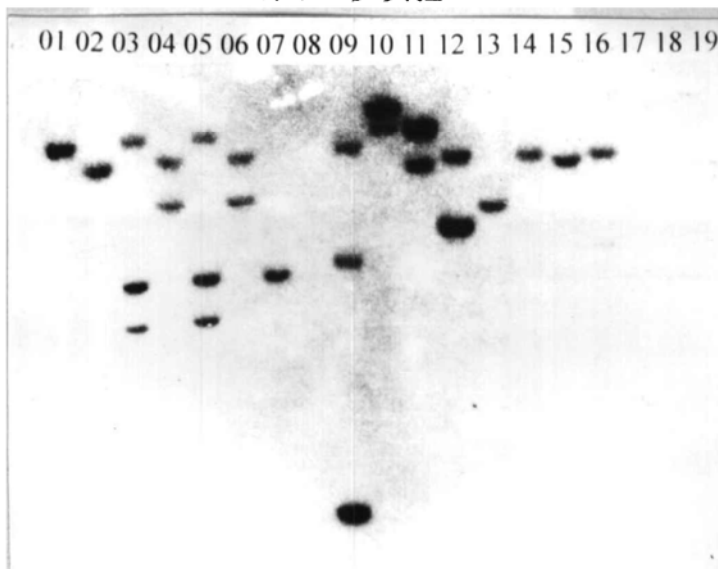


图 11: 以 3'端序列为探针的 southern 杂交结果 (序号同图 8)

八、 将 124A22 以 HindIII 酶切后的 6 条杂交阳性条带 1、2、3、6、7、10 从凝胶中回收, 这 6 条插入片段的大小分别约为 9kb、6kb、4kb、2.5kb、1kb、0.5kb, 克隆进以 HindIII 酶切后的 pBS 载体中, 将含有上述片段的克隆定名

为 pBS-MDR101、pBS-MDR102、pBS-MDR103、pBS-MDR106、pBS-MDR107、pBS-MDR110。

电穿孔感受态细胞的转化效率约 10^8 - 10^9 cfu/ μ g, 去磷酸化后自连效率低于 1%。

九、 亚克隆的单一位点的限制性内切酶

pBS-MDR101	BamHI*	EcoRV
pBS-MDR102	NotI*	BamHI
pBS-MDR103	XbaI*	EcoRV
pBS-MDR106	XbaI*	EcoRV
pBS-MDR107	XbaI*	EcoRV

其中*者为第一次酶切所选用的酶。

十、 5 种亚克隆经 ExoIII、S1 核酸酶作用后 (图 12、13、14), 各得到一系列彼此之间相差 450bp 左右的次级亚克隆, 此结果符合 ExoIII 在 37°C 时每分钟降解速率为 450bp 文献报道。

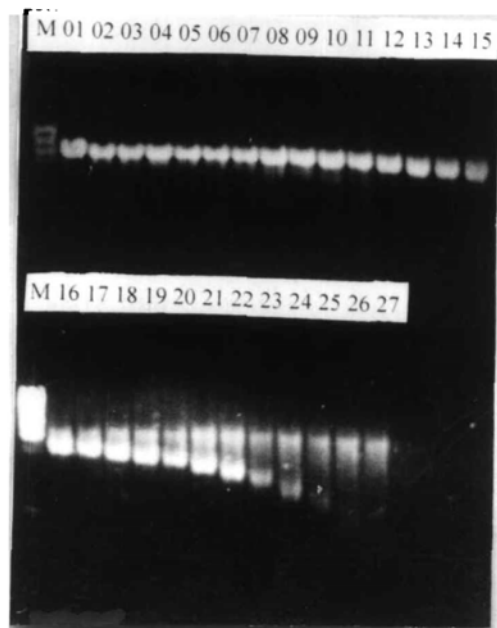


图 12: pBS-MDR101 的 ExoIII 酶切图 (M: λ DNA/BstEII)

十一 各次级亚克隆的转化效率一般为 10^2 - 10^3 转化子/20ngDNA。

十二 96 孔板中的各次级亚克隆经过快速裂解后电泳鉴定 (图 15), 根据各次级亚克隆的超螺旋状态在凝胶中的位置, 选取了 53 个次级亚克隆 (图 16) 基本覆盖了 5 种亚克隆的全长序列, 将各次级亚克隆分别命名为:

pBS-MDR10101- --pBS-MDR10121、 pBS-MDR10201- --pBSMDR10213、
 pBS-MDR10301- --pBS-MDR10312、 pBS-MDR10601- --pBSMDR10605、
 pBS-MDR10701- --pBS-MDR10702

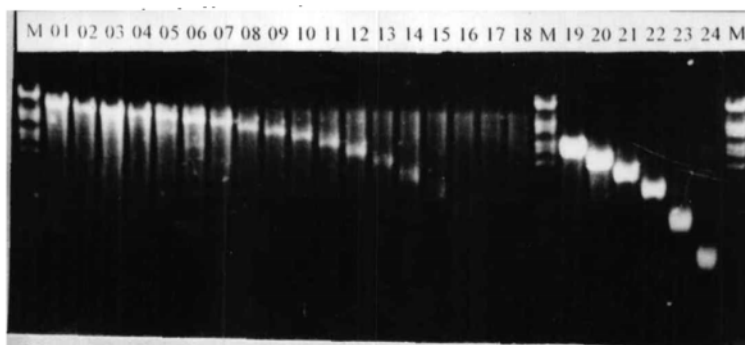


图 13: pBS-MDR102 (左) 和 pBS-MDR107 (右) 的 ExoIII 酶切图

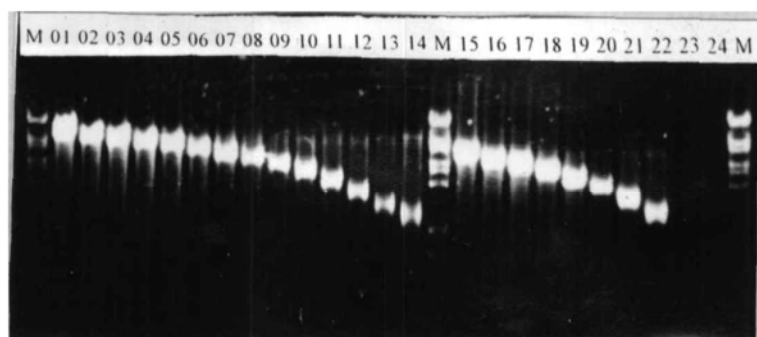


图 14: pBS-MDR103 (左) 和 pBS-MDR106 (右) 的 ExoIII 酶切图

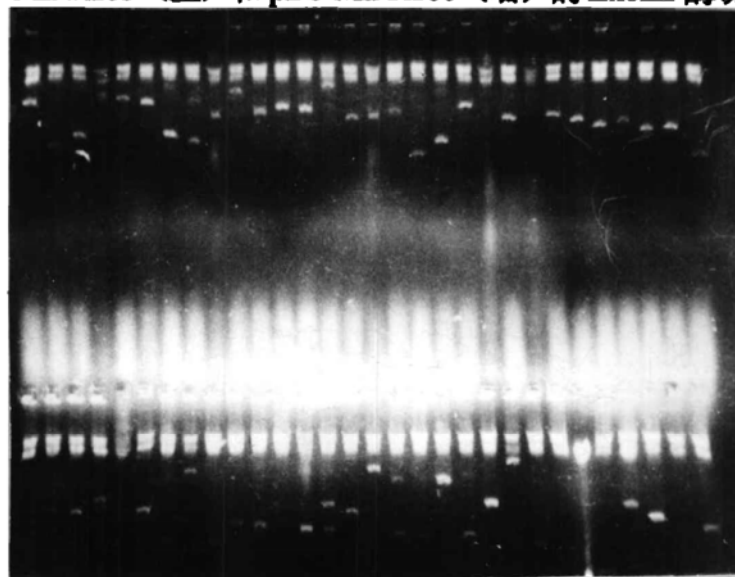


图 15: 快速裂解法验证质粒大小图

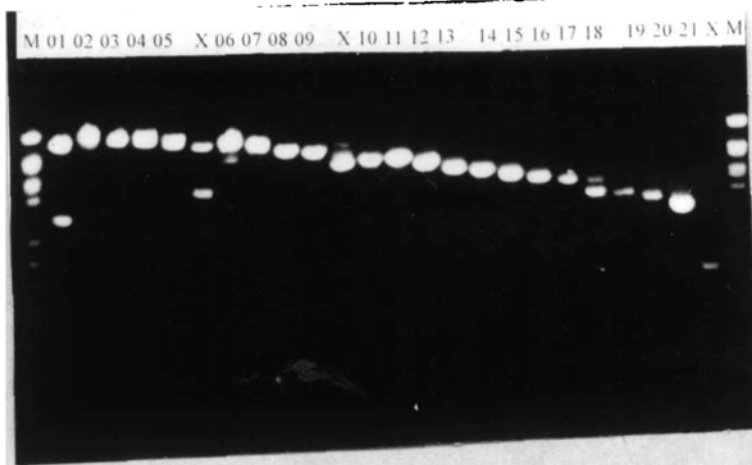


图 16: pBS-MDR101 的 21 个次级亚克隆的梯度图

M:λDNA/EstEII; x: 不符合梯度要求的次级亚克隆

十三 以 T_3 或 M13R 为引物, 通过 ABI377 全自动测序仪对上述 53 个次级亚克隆进行测序, 测序部分结果 (pBS-MDR10118) 如下:

十四 将上述测序结果送至数据库中, 比较同源性(所用的软件为 blastx 和 blastn), 结果表明, 在核苷酸水平上, 河豚鱼基因组序列与人的基因组序列有着很大的差异, 仅发现 pBS-MDR10118 与人在核苷酸水平有较高的同源性 (如下):

```

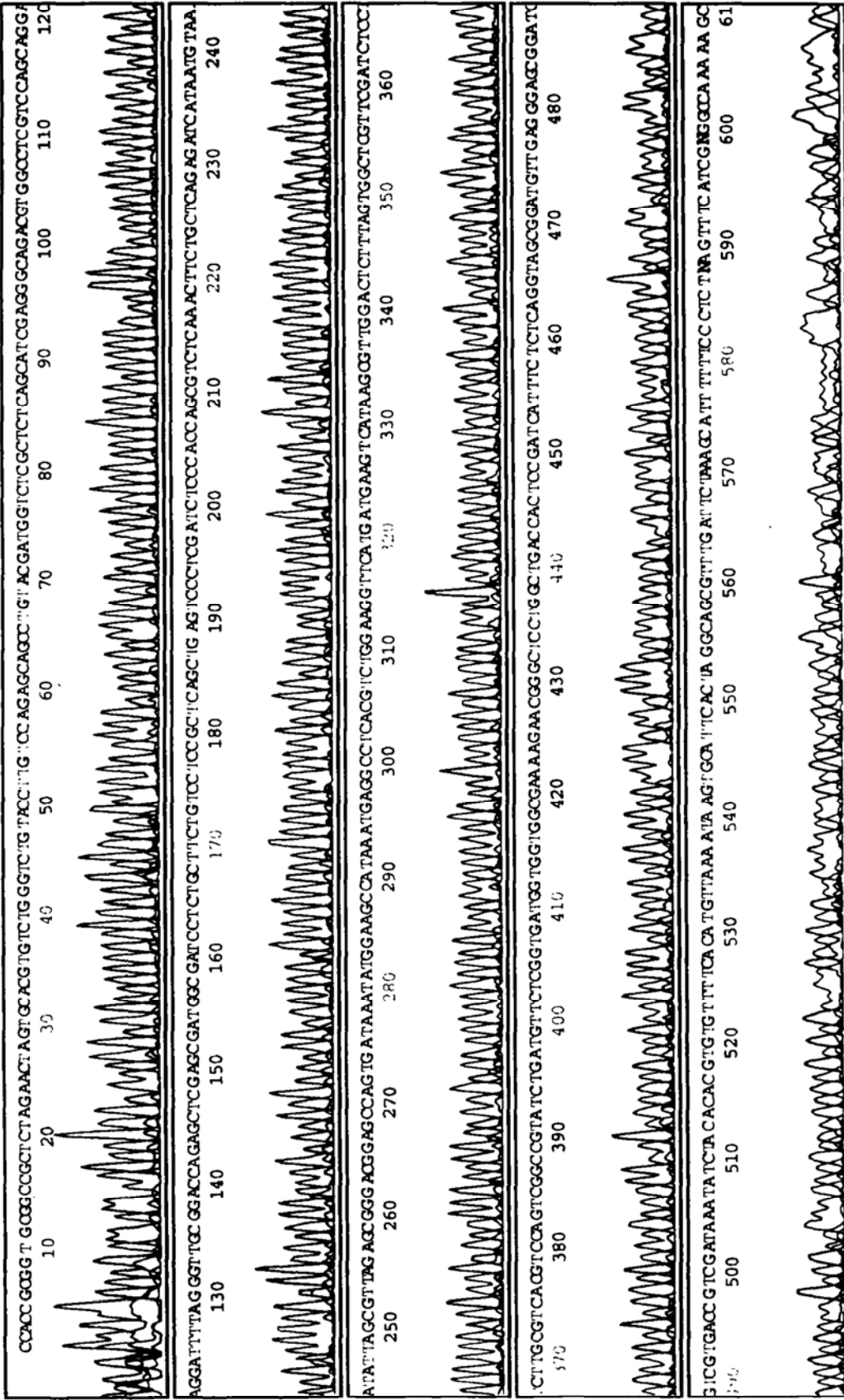
gb|M23234|HUMMDR3 Human membrane glycoprotein P (mdr3) mRNA, complete cds.
    Length = 3924
    Score = 89.7 bits (45), Expect = 5e-16
    Identities = 105/125 (84%), Positives = 105/125 (84%)
Query: 100  gacgtggcctcgtccagcaggaggattttaggggtgcggaccagagctcgagcggatggcg 159
            ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct: 1714 gacgtggcctcatccagcagaaggatcttgggggtgcgaaccagggcagctgcaatggcg 1655
Query: 160  atcctctgcttctgtcctccgctcagctgagtcctcgcgatctcccaccagcgtctcaaac 219
            ||| | | | | | | | | | | | | | | | | | | | | | | |
Sbjct: 1654 atcctctgcttctgcccaccactcagctgggcccctctctctccaaccagggtgtcaaat 1595
Query: 220  ttctg 224
            ||| |
Sbjct: 1594 ttctg 1590

gb|M29433|HUMMDR1A12 Human P-glycoprotein (MDR1) gene, exon 14.

```

Model 377
Version 3.0
ABI100
Version 3.0
18-18
18
Lane 18
Signal C:53 A:69 G:172 T:102
DP 4%Ac(T3)
DP & DT Matrix set A
Points 1204 to 10170 Base 1: 1204

Page 1 of 2
Sat, Jan 24, 1998 10:23 AM
Fri, Jan 23, 1998 8:24 PM
Spacing: 11.14(11.14)



```
Length = 215
Score = 87.7 bits (44), Expect = 2e-15
Identities = 104/124 (83%), Positives = 104/124 (83%)
Query: 100 gacgtggcctcgtccagcaggaggattttagggttgCGGaccagagctcgagcgatggcg 159
      |||
Sbjct: 139 gacgtggcctcatccagcaggaggatcttggggttgCGaaccagggcacgtgcaatggcg 80
Query: 160 atcctctgcttctgtcctccgctcagctgagtcctcgcgatctcccaccagcgtctcaaac 219
      |||
Sbjct: 79 atcctctgcttctgcccaccactcaactgggcccctctctctccaaccagggtgtcaaat 20
Query: 220 ttct 223
      |||
Sbjct: 19 ttct 16
```

```
gb|M14758|HUMMDR1 Human P-glycoprotein (MDR1) mRNA, complete cds.
>gi|2471830|gb|I49610|I49610 Sequence 2 from patent US
5641508
Length = 4646
```

```
Score = 85.7 bits (43), Expect = 8e-15
Identities = 100/119 (84%), Positives = 100/119 (84%)
Query: 100 gacgtggcctcgtccagcaggaggattttagggttgCGGaccagagctcgagcgatggcg 159
      |||
Sbjct: 2100 gacgtggcctcatccagcaggaggatcttggggttgCGaaccagggcacgtgcaatggcg 2041
Query: 160 atcctctgcttctgtcctccgctcagctgagtcctcgcgatctcccaccagcgtctcaaa 218
      |||
Sbjct: 2040 atcctctgcttctgcccaccactcaactgggcccctctctctccaaccagggtgtcaaa 1982
```

```
gb|AF016535|AF016535 Homo sapiens P-glycoprotein (mdr1) mRNA, complete cds
Length = 4192
Score = 85.7 bits (43), Expect = 8e-15
Identities = 100/119 (84%), Positives = 100/119 (84%)
Query: 100 gacgtggcctcgtccagcaggaggattttagggttgCGGaccagagctcgagcgatggcg 159
      |||
Sbjct: 1792 gacgtggcctcatccagcaggaggatcttggggttgCGaaccagggcacgtgcaatggcg 1733
Query: 160 atcctctgcttctgtcctccgctcagctgagtcctcgcgatctcccaccagcgtctcaaa 218
      |||
```

Sbjct: 1732 atcctctgcttctgccaccactcaactgggccctctctccaaccagggtgtcaaa 1674

但在氨基酸水平上,发现大部分上述次级亚克隆均与人或其他生物的多药耐药基因或 ATP 结合蛋白有着较强的同源性,尤其是 pBS-MDR101 的各次级亚克隆与人的 MDR1 基因的同源性遵循着肯定的规律性:

sp|P08183|MDR1_HUMAN MULTIDRUG RESISTANCE PROTEIN 1 (P-GLYCOPROTEIN 1)

>gi|1070659|pir||DVHUI multidrug resistance protein 1 -
human >gi|386862 (M29447) P glycoprotein [Homo sapiens]
Length = 1280

>No. pBS-MDR10118

Score = 189 bits (474), Expect = 2e-47

Query: 509 IFIDGHDIRSLNIRYLREMIGVVSQEPVLFATTITENIRYGRDVTQEEIERATKESNAY 330

+ +DG DIR++N+R+LRE+IGVVSQEPVLFATTI ENIRYGR +VT +EIE+A KE+NAY

Sbjct: 451 VSDGQDIRTINVRFLREIIGVVSQEPVLFATTIAENIRYGRENTMDEIEKAVKEANAY 510

Query: 329 DFIMNLPDVRPHLWLPYLSLAPSRSNANIYIMISEQKFETLVGDRGTQLSGGQKQRIAIA 150

DFIM LP KF+TLVG+RG QLSGGQKQRIAIA

Sbjct: 511 DFIMKLP-----HKFDTLVGERGAQLSGGQKQRIAIA 542

Query: 149 RALVRNPKILLLDEATSALDAESETIVQAALDKVQTQTRALVL 21

RALVRNPKILLLDEATSALD ESE +VQ ALDK + +V+

Sbjct: 543 RALVRNPKILLLDEATSALDTESEAVVQVALDKARKGRITIVI 585

Score = 146 bits (364), Expect = 2e-34

Query: 509 IFIDGHDIRSLNIRYLREMIGVVSQEPVLFATTITENIRYGRD---VTQEEIERATKESN 336

+ +DG +I+ LN+++LR +G+VSQEP+LF +I ENI YG V+QEEI RA KE+N

Sbjct: 1094 VLLDGKEIKRLNVQWLAHLGIVSQEPILFDCSIAENIAYGDNSRVVSQEEIVRAAKEAN 1153

Query: 335 AYDFIMNLPDVRPHLWLPYLSLAPSRSNANIYIMISEQKFETLVGDRGTQLSGGQKQRIA 156

+ FI +LP+ K+ T VGD+GTQLSGGQKQRIA

Sbjct: 1154 IHAFIESLPN-----KYSTKVGDKGTQLSGGQKQRIA 1185

Query: 155 IARALVRNPKILLLDEATSALDAESETIVQAALDKVQTQTRALVL 21

IARALVR P ILLLDEATSALD ESE +VQ ALDK + +V+

Sbjct: 1186 IARALVRQPHILLLDEATSALDTESEKVVQEAALDKAREGRTICIVI 1230

>No. pBS-MDR10116

Score = 81.5 bits (198), Expect = 5e-15

Query: 444 QVRLGRTTIVIAHRLSTIRNADIIAGFSNGEIVEGQTHSQLMEIKGVYHGLVTMQVS 274

+ R GRTTIVIAHRLST+RNAD+IAGF +G IVE+G H +LM+ KG+Y LVTMQ +

Sbjct: 575 KARKGRTTIVIAHRLSTVRNADVIAGFDDGVIVEKGNHDELMKEKGIYFKLVTMQTA 631

Score = 69.5 bits (167), Expect = 2e-11

Query: 444 QVRLGRITIVIAHRLSTIRNADIAGFSNGEIVEQGTHSQLMEIKGVYHGLVTMQ 280

+ R GRT IVIAHRLSTI+NAD+I F NG + E GTH QL+ KG+Y +V++Q

Sbjct: 1220 KAREGRTCIVIAHRLSTIQNADLIVVFCNGRVKEHGTHQQLLAQKGIYFSMVSQ 1274

>No. pBS-MDR10114

Score = 74.5 bits (180), Expect = 7e-13

Query: 304 DENVPPVSFFKVMRYNVSEWPYILVGTICAMINGAMQPVSIIIFTEIIMV 155

DE++PPVSF+++M+ N++EWPY +VG CA+ING +QP F+IIF++II V

Sbjct: 689 DESIPPVSFWRIMKLNLTWPYFVVGVRCAIINGLQPAFAIIFSIIIGV 738

>No. pBS-MDR10113

Score = 34.8 bits (78), Expect = 0.61

Query: 424 REKDKEIIREKSSFFCILFAVMGVVFTLTMFLQ 326

R D E R+ S+ F +LF +G+++F+T FLQ

Sbjct: 741 RIDDPETKRQNSNLFSLFLALGIISFITFFLQ 773

>No. pBS-MDR10110

Score = 46.5 bits (108), Expect(2) = 4e-20

Query: 442 KDLIWYDXPKNTVGALTTRLAADAHVQG 356

+D+ W+D PKNT GALTTRLA DAA V+G

Sbjct: 799 QDVSWFDDPKNTTGALTTRLANDAAQVKG 827

Score = 72.6 bits (175), Expect(2) = 4e-20

Query: 282 AAGVRLAVMTQNFANLGTSTIIISFYVGWEXXXXXXXXXXXXXXXXXXXAEVKLLTGHAEDK 103

A G RLAV+TQN ANLGT IIISF+YGW+ E+K+L+G A +DK

Sbjct: 828 AIGSRLAVITQNIANLGTGIIISFYGWQLTLLLAIVPIIAIAGVVEMKMLSGQALKDK 887

Query: 102 KELEMAGKV 76

KELE AGK+

Sbjct: 888 KELEGAGKI 896

>No. pBS-MDR10108

Score = 33.6 bits (75), Expect = 1.4

Query: 177 SVVMTMLYGAMAVGEANTYAPNFAKAKISASHLTMLINRQPAIDNLSE 34

+V ++L GA +VG+A+ FA A+ +A + +I+ +P+ID+ S+

Sbjct: 333 TVFFSVLIGAFSVGQASPSIEAFANARGAAYEIFKIIDNKPSIDSYK 380

Score = 56.6 bits (134), Expect = 2e-07

Query: 174 VVMTMLYGAMAVGEANTYAPNFAKAKISASHLTMLINRQPAIDNLSEE 31

V +++GAMAVG+ +++AP++AKAKISA+H+ M+I + P ID+ S E

Sbjct: 977 VFSAVVFGAMAVGQVSSFPDYAKAKISAHHIIMIIEKTPLIDSYSTE 1024

>No. pBS-MDR10107

Score = 54.7 bits (129), Expect = 7e-07

Query: 315 IGEANTYAPNFAKAKISASHLTMLINRQPAIDNLSEXEARLVNDTXPHTTVDALK*QQAX 136

+G+ +++AP++AKAKISA+H+ M+I + P ID+ S P+T

Sbjct: 988 VGQVSSFPDYAKAKISAHHIIMIIEKTPLIDSYSTEGLMPNT 1030

Query: 135 *HTVLLCRLQEKYDGNVLFEDXKFNYPSPRPDV 40

+GNV F + FNYP+RPD+

Sbjct: 1031 LEGNVTFGVVFVFNYPTRPDI 1050

>No. pBS-MDR10106

Score = 106 bits (263), Expect = 1e-22

Query: 489 DGNVLFEDVKFNYPSPRPDVPVQLGLNLEVQKGETLALVGSSGCGKSTTIQLLERFYDPRE 310

+GNV F +V FNYP+RPD+PVLQGL+LEV+KG+TLALVGSSGCGKST +QLLERFYDP

Sbjct: 1032 EGNVTFGVVFVFNYPTRPDI PVLQGLSLEVKKGQTLALVGSSGCGKSTVQQLLERFYDPLA 1091

Query: 309 GRVV 298

G+V+

Sbjct: 1092 GKVL 1095

Score = 91.3 bits (223), Expect = 6e-18

Query: 486 GNVLFEDVKFNYPSPRPDVPVQLGLNLEVQKGETLALVGSSGCGKSTTIQLLERFYDPREG 307

GN+ F +V F+YPSR +V +L+GLNL+VQ G+T+ALVG+SGCGKST+QL++R YDP EG

Sbjct: 390 GNLEFRNVHFSYPSRKEVKILKGLNLKVQSGQTVALVGNSSGCGKSTTVQLMQRLYDPTEG 449

Query: 306 RV 301

V

Sbjct: 450 MV 451

>No. pBS-MDR10105

Score = 153 bits (384), Expect = 8e-37

Query: 532 AAHDHLHHIQTXXSEQRYDTQAGDKGTQLSGGQKQORVAIARAIIRNPKLLLLDEATSALDT 353

A ++H +Y T+ GDKGTQLSGGQKQOR+AIARA++R P +LLLDEATSALDT

Sbjct: 1149 AKEANIHFIESLPNKYSTKVGDKGTQLSGGQKQORAIARALVRQPHILLDEATSALDT 1208

Query: 352 ESEKRLYSH*CYVVHKXAAQHCTSCVQVVQEALDQARKGRTCIVVAHRLSTIQNAD CIA 173
 ESEKV VQEALD+AR+GRTCIV+AHRLSTIQNAD I
 Sbjct: 1209 ESEKV VQEALDKAREGRTCIVIAHRLSTIQNADLIV 1244
 Query: 172 VFQGGVVVEKGTHQQLIAKKG 110
 VFQ G V E GTHQQL+A+KG
 Sbjct: 1245 VFQNGRVKEHGTHQQLAOKG 1265
 Score = 139 bits (346), Expect = 2e-32
 Query: 487 RYDTQAGDKGTQLSGGQKQ RVAIARAI IRNPKLLLLDEATSALDTESEKRLYSH*CYVV 308
 ++DT G++G QLSGGQKQ+AIARA++RNPK+LLLDEATSALDTESE
 Sbjct: 519 KFDTLVGERGAQLSGGQKQRIAIARALVRNPKILLDEATSALDTESE 566
 Query: 307 HKXAAQHCTSCVQVVQEALDQARKGRTCIVVAHRLSTIQNAD CIA VFQGGVVVEKGTHQQL 128
 VVQ ALD+ARKGRT IV+AHRLST++NAD IA F GV+VEKG H +
 Sbjct: 567 AVVQVALDKARKGRTTIVIAHRLSTVRNADVIAGFDDGVIVEKGNHDE 614
 Query: 127 MIAKKG 110
 ++ +KG
 Sbjct: 615 LMKEKG 620

统计结果如下:

次级亚克隆序号	碱基位置	对应的人 MDR1 氨基酸位置
No. pBS-MDR10118	509-21	451-585 (1094-1230)
No. pBS-MDR10116	444-274	575-631 (1220-1274)
No. pBS-MDR10114	304-155	689-738
NO. pBS-MDR10113	424-326	741-773
NO. pBS-MDR10110	442-356	799-827
	282-76	828-896
No. pBS-MDR10108	177-31	977-1024 (333-380)
No. pBS-MDR10107	315-40	988-1050
No. pBS-MDR10106	489-298	1032-1095 (390-451)
NO. pBS-MDR10105	532-110	1149-1265 (519-620)

讨 论

河豚鱼作为一种新型模式生物,有其自身独特的特点:①“密集”的基因组序列,其全长序列只有人类基因组的 1/7—1/8;②基因间序列小,重复序列少;③基因内含子小,但外显子的数目与人基本一致;④与人类的进化距离较近,属最原始的脊椎动物。利用河豚鱼的上述特点可为人类基因组计划提供一全新的参照系,在序列测定、比较基因组学、致病基因的克隆、基因的结构功能和表达调控等方面的研究中发挥重要的作用。自 1993 年 Brenner 教授在 *Nature* 上发表第一篇有关河豚鱼作为模式生物的论文以来,英、德、美等国已将其列为人类基因组计划中的一项重要内容而加以大力资助,全球已有相当的实验室将其列为重要研究对象。但限于以前对于河豚鱼基因组较肤浅的认识和发展时间较短,目前,大多数关于河豚鱼基因组计划的研究还限于两个方面:①理论的验证,即验证河豚鱼是否是最佳的模式生物之一,其基因组数目是否与人类相似,其各基因结构,即外显子、内含子数目及剪接等是否与人的同源基因一致,基因间即基因组在进化上是否保守;②对获得的与人同源的河豚鱼基因进行比较基因组学研究。随着理论的进一步完善,下一步将利用这些同源基因来研究基因的结构功能和表达调控,利用同线保守理论来协助克隆人类疾病相关基因。

本文的主要目的亦即是先验证河豚鱼作为模式生物理论,选择人的多药耐药基因全长 cDNA 作为探针筛查河豚鱼 cosmid 文库。多药耐药机制是肿瘤化疗中的一大难题,多药耐药基因的转基因疗法将有效的缓解这一矛盾,虽然目前多药耐药基因已在结构、功能和表达、调控方面得到了很彻底的认识,但我们并不能说我们已经获得了这个基因的全部信息,比如内含子的功能、全长基因与 cDNA 之间的表达差异等等,换句话说,如果我们能获得某些位于内含子中或位于别处的多药耐药基因的调控序列如增强子,将这些调控序列与人的多药耐药基因的全长 cDNA 重新组装后再进行转基因治疗,那将会进一步扩大基因治疗的效果。同时河豚鱼作为一种携带有剧毒毒素—河豚毒素的特殊生物,必定有其独特的耐药机制,如果能从多药耐药基因方面找到有关河豚鱼的耐药机制,那将给毒理学的研究提供一些新的线索。既然多药耐药基因在临床与基础方面都具有如此重要的意义和作用,故本文选择多药耐药基因作为研究的对象。

人的多药耐药基因有二种亚型,即 MDR1 和 MDR3。一般只有 MDR1 才在大部分组织中表达而发挥多药耐药功能,而在小鼠中多药耐药基因有三种亚型。多药耐药基因作为看家基因之一,从原始的大肠杆菌到酵母,从无脊椎动物到哺乳动物,许多生物中都存在此基因或同源基因。多药耐药基因作为 ATP 结合蛋白家族中的一员,和其它的 ATP 结合蛋白一样,其编码产物为一种跨膜蛋白,依靠

ATP 提供能量而将细胞内的一些物质“泵”出细胞外,其作用底物多种多样,而不象其它蛋白只有一种或几种底物。该基因在细胞的“自我防御”功能上起着重要的作用。从理论上推测河豚鱼也应有这种基因。

探针选用的是人的 MDR1 全长 cDNA,其理由是:河豚鱼与人的进化距离是 450 万年,即使河豚鱼存在这种基因,且该基因在功能上与人保守,但该基因在核苷酸水平上也将会由于突变事件的发生而与人存在较大的差异,但一般在整个基因序列中,功能特异性区域的保守性应相对较强,在不知道哪段序列为功能特异性区域或功能特异性区域较难获得的情况下,采取基因的全长 cDNA 作为探针是较为理想的方法。当然,目前国际上其它实验室往往采取兼并探针,即根据功能保守区的两侧序列设计引物,以 PCR 产物作为探针来筛查河豚鱼文库。

由于河豚鱼 cosmid 文库杂交膜是从英国剑桥大学人类基因组计划中心购买,价格不菲,且杂交系统是异种杂交,所以本文首先采取预实验。预实验的目的之一是验证河豚鱼基因组是否肯定存在多药耐药基因,也就是验证河豚鱼基因组中是否存在与人类多药耐药基因全长 cDNA 同源的序列;二是摸索异种杂交的条件。一般采取 church 杂交系统进行同种杂交时,杂交温度为 65°C,预杂交 5hr,杂交 12hr,预实验中发现,当杂交温度为 65°C 时,杂交信号较弱甚至有些条带消失,而杂交温度为 50°C 时,杂交背景较高,55°C 和 60°C 时杂交条件较合适,考虑到 55°C 杂交时获得的假阳性信号会多于 60°C,故选择 60°C 作为异种杂交温度。

河豚鱼 cosmid 文库杂交时由于是菌落杂交,杂交环境与 southern 杂交不太一样,为尽量降低细菌碎片和其它一些因素对杂交结果的影响,本文采取延长预杂交时间至 12hr,加大封闭试剂量,预杂交液中含有超声打碎的鲑精 DNA 200—300μg/ml,同时含有 100μg/ml 超声打碎的 cosmid 载体(该载体为本室构建河豚鱼 cosmid 文库时所用,与英国的河豚鱼 cosmid 文库的载体不一样,但序列大致相似)。

在洗膜的过程中,曾出现膜的杂交背景信号很高,依靠一般的洗脱方法很难去除,作者曾尝试过将洗脱液 II 的洗脱温度提高到 80°C,经过 30min 左右才将膜上的背景信号降低至正常范围,当然此时不再可能看到杂交阳性信号。分析原因可能是由于杂交膜处理的过程中将细菌碎片去除得不太充分,以致非特异吸附的能力很强,此种现象也曾出现在本室自制的河豚鱼 BAC 文库的杂交膜的杂交过程中,在杂交之前仔细处理过的杂交膜(指去除细菌碎片)容易洗脱,而经一般处理后的膜杂交后,杂交背景很强且很难去除。

为了最大限度的延长该河豚鱼杂交膜的使用寿命,杂交时采取的是三种探针的混合物:人的 MDR1 cDNA、人的 α -TNF cDNA 和人 5 号染色体短臂端经过 exon-trapping 获得的一段 300bp 左右的外显子,在结果中所获得的 44 个候选克隆

为三种探针的混合杂交结果。

将候选克隆酶切、转膜后以人的 MDR1 cDNA 为探针杂交, 获得 10 个阳性克隆, 在这 10 个阳性克隆中, 酶切结果和杂交结果显示: 彼此之间有一些条带的大小是一致的, 这从侧面可以反映出这 10 个克隆是从河豚鱼基因组同一段序列中经过不完全酶切而获得的大小不一的片段, 而正因为这 10 个克隆均与人 MDR1 有着较强的同源性, 说明河豚鱼基因组应存在着与人多药耐药基因同源的基因。

以 5' 序列为探针获得的杂交信号很弱, 阳性条带数很少, 对此现象的解释是: 5' 端序列主要是一些调控序列而非功能特异性区域, 故在不同种的同源基因之间的保守性很差, 所以杂交信号很弱。

以 3' 端序列为探针获得的杂交信号很强, 且条带数很多, 对此现象的解释是: 一、3' 端序列在不同物种的同源基因之间的保守性较高; 二、根据小鼠和其它生物的多药耐药基因结构分析, 3' 端序列和 5' 端半侧的 3' 端有很强的同源性, 即基因内不同部位有着很强的同源性, 河豚鱼的多药耐药基因也可能有着上述特性; 三、河豚鱼基因组中存在多药耐药基因家族。

MDR1 杂交后的 6 条阳性条带的总长度约为 23kb, 如果这为一个基因的长度, 也只有人的 MDR1 的 1/5 或更小, 如果这是一个基因家族的长度, 那河豚鱼的多药耐药基因只有人的 1/10 或更小。

测序是分子生物学最常见的技术之一: 目前的测序技术一般满足每一次只能测出 500bp 左右的要求, 这对于测定大片段序列来说, 建立各种亚克隆成为摆脱这一瓶颈的最有效方法之一。传统上, 可采用反复酶切, 以获得一个个适合测序要求的酶切片段, 再亚克隆。此法对于 2-3kb 的序列较有效, 但对于大片段序列将面临各片段间如何相互重叠的问题, 同时还需要许多种类不同的限制性内切酶。国外最常见的方法是通过超声打碎即鸟枪法(shotgun)来构建满足测序要求的亚克隆。该法简捷高效, 但耗资巨大, 且需要复杂的软件系统, 显然不适合我国的科研实情。对于一般不太长的序列(3kb 左右), 我们通常的方法是通过引物步移来获得全长序列, 该法的缺点是耗时长, 不适合测更长片段的序列, 同时由于必须合成引物使成本较高。

外切酶 III 具有从非 4 个碱基 3' 凸端逐个降解核苷酸, 末端为 α - ^{35}S -dNTP 将对其具有屏蔽作用, 及在一定的温度条件下降解速率恒定的两大特点。其最适反应温度为 22-46°C, 且每升高 6°C, 降解速率增加一倍^[1]。一般 37°C 时每分钟降解 450bp^[2]。自 1984 年 Steven Henikoff^[3] 建立起此方法后, 该法曾一度风靡生物学界。但由于该法具有以下几个缺点而使其应用过程变得十分麻烦: 1. 必须保证 DNA 的超螺旋化状态达 80% 以上; 2. 需找出一靠近载体引物侧且产生 4 个碱基 3' 凸端的限制性内切酶来消化, 或以 α - ^{35}S dNTP 补平第一次酶切后产生的末

端；3. 验证各单克隆是否含有所需要大小的插入片段时，如果通过质粒提取后再电泳鉴定，工作量非常大。国外文献报道可通过对单克隆的直接裂解法来快速验证^[2]，但往往由于国内实验室各种条件及试剂的限制，单克隆直接裂解法获得的质粒量很少，凝胶中几乎看不出；4. 次级克隆的挑选和鉴定十分繁琐，因为每个平板中有 1/3 甚至更多的克隆不含有其对应时间所应有大小的片段。作者在实验中发现可通过下述方法来克服上述困难：1. 超螺旋化 DNA 的获得可以不通过 CsCl 梯度离心而改从凝胶中回收超螺旋片段；2. 将单克隆挑入 96 孔板中增殖，辅以多道移液器等，再通过裂解法，可一人一天鉴定数百个克隆；3. 挑选次级亚克隆时不必拘泥于非得从每一个平板中来挑选其所对应时间段应获得的正确大小的片段，而可以从一个平板中挑选另一个时间段所对应的克隆，只要其大小符合前后差异 450bp 左右的要求即可。

连接、转化效率也是实验中需注意的问题。一般我们使用化学法转化时的感受态细胞的效率要求达到 10^7 cfu/ μ g 才能获得数十至数百个克隆。虽然现在我们有许多种方法可以用来制备高效率的感受态细胞，但由于各种因素的影响，实际上很难恒定地得到 $>10^7$ cfu/ μ g 的感受态。连接效率也值得注意，该法是一平端自连的过程，国外文献报道只须常温连接 2hr 即可，但实际操作中发现这样的连接效果较差。为减少连接和转化对结果的影响，本文采取以下办法：(1) 16°C 连接过夜且用较大量的 T_4 连接酶；(2) 电穿孔法转化，必要时将连接产物纯化后全部转化(因连接产物的离子浓度很大，未纯化时只能取 1-2 μ g 用来转化)。

另外，实验中发现：由于一般实验室电泳技术的限制，尚不能有效区分 10kb 以上的克隆条带之间是否只是 450bp 左右的大小差异。这样会经常碰到前后两个克隆在凝胶上似乎差别很小但序列之间却没有重叠部分，这种情况下一般可通过一次引物步移来填补空缺。所以，在利用 ExoIII 建立次级亚克隆测序系统时，首先尽量通过限制性内切酶将目的片段控制在 8Kb 以下。

当然，也可以通过 ExoIII 与聚合酶延伸^[4]和特殊载体(double-origin Vector)^[5]等方法的结合来获取大片段序列的次级亚克隆。

pBS-MDR101 的 21 个次级亚克隆之间并不是完全都能互相重叠，尤其是靠前的几个次级亚克隆，产生这现象的原因前面已有交代，但是通过 ExoIII 而获得的 pBS-MDR101 的次级亚克隆之间的 blastx 结果呈现出肯定的连续性，这说明 ExoIII 确实是一种有效的亚克隆方法，虽然本文并未获得一完整的彼此之间都相互重叠的次级亚克隆。

总之，利用 ExoIII 来建立次级亚克隆测序系统在国内仍不失为一快速、高效、经济的方法。

测序结果表明：河豚鱼确实含有与人同源的多药耐药基因，pBS-MDR101 的次级亚克隆序列中可以看出明显的含有一条完整的多药耐药基因，由于资金和设备的原因，只对其它的亚克隆的部分次级亚克隆进行测序，结果没有看出有上述

明显的趋势，但大部分都与人或其它生物的多药耐药基因有同源性，据此，我们可以初步认定：1 河豚鱼中肯定含有与人同源的多药耐药基因；2 河豚鱼中至少含有 2 个多药耐药基因。对基因的进一步分析，可通过测定全长序列后进行。这正是论文的第二部分。

序列比较结果发现：在核苷酸水平上，pBS-MDR101 的 21 个亚克隆中仅 pBS-MDR10118 与人多药耐药基因有着较强的同源性，而其它次级亚克隆均未发现，但在氨基酸水平上，大部分次级亚克隆均与人的多药耐药基因有着较强的同源性，这表明：1 对于异种杂交，采取全长 cDNA 作为探针是正确和可靠的，但不能避免大量假阴性结果，也同样会产生一些假阳性结果，但通过反复杂交验证，可有效的将假阳性结果排除；2 基因在进化的过程中，虽然存在着较多的突变事件，但由于兼并密码子的存在，使得基因的功能得以保存。

第二部分

河豚鱼 MDR 基因组全长序列分析及其邻座基因同线分析

实验策略

河豚鱼 *mdr* cosmid 阳性克隆

- | 超声波打碎
- | 平均大小为 500bp
- | 与测序载体连接

shotgun 法测序亚克隆

- | 测序、序列重叠分析

河豚鱼 *mdr* cosmid 全长序列 40kb

- | Grail, sspn, Genie, blastn, blastx,
- | blastp, dotplot, seqlab
- | 等软件分析系统

河豚鱼 *mdr* cosmid 含有 2 个完整的多药耐药基因和 4 个其他基因

I. Cosmid 124A22 全长序列的测定

通过超声打碎法获得 cosmid 124A22 的测序随机重叠克隆 M13 文库，随机挑选 400 个克隆，提取质粒以 ABI377 测序，所测序列经过计算机自动处理，搭好重叠群，去除载体序列，最后得到两个重叠群 contig1 和 contig2，结果如下：

>Contig1

```

ATACTGGAAATGCTATATTGCAACCTCACACGCGACTCACAGGTTTTTGACCAGAAAGGAATAAAAAACAAAAACTATA
CACAACTTAAATGACCATGACGCTACATAAGCTCTGAACGGATGTATGTCGATTACTGGGGGCAATATGACACAACTC
ACATCCAGCAGCATTTCAAATGAGCAGGACCACAACCAGCTGTGCACAAACATCTGTGGGAAACATTCAGCTAAAACTCC
ACAAAGTAGATGCGCTCACGTACAATTGCACTTAGCTGTAATCATCACNTCGGTGCACAGTACAAGTTGATGCTGATTAGAG
AAGATTTGGGATGGCAGAAGCTGGTAAAGGTGTCCCAAGTCCGAGTGTGTGACGCGCCATTGCTGAGGAAAGGTTACTGC
ACGCCTTCCTTACAGCTTTGGAGAAAAAGGTGGACTGAGGTAGAAGGGAAAAAGCCAAGAAGGCAAAAGAAGGCAAGAAA
TAAGGAGTGTATCTGTTCCACAAGCAAGCTGCAGGGTTGGTGTGTTCAAAGGCACTTGGAAACATGAGGAAAAGGTTGGAA
AAAAACCAAAAAACACACAAAATATTTGGTTGAGGTCAAAATGTTGCATGGTGACTTGGTTTAGCTAGGTAAGGCGGG
AGAGGAAGAAAAGCCCTTAACATACTCAAACAAATAGCCTCCTTCTGCGGGTGGAAACACTTCATGTTGGCAATTAATTCGG
AGTAAATAAAATTTACAAAAT

```

contig2

```

TTANGATTGTTTTCCATGGAGCTGCCATGACCATTTACACCATATACATACTGTCTCTGAGCAGAGATACGACACTCAGGC
TGGTGATAAGGGAACACAGCTGTCAGGGGGCCAGAAGCAGCGTGTGCCATAGCCCGAGCCATCATCCGCAACCCCAACTG

```

TTGCTCCTGGACGAGGCCACGTCTGCGCTCGACACTGAGAGTGAGAAGGTGAGACTTTATTACATTAGTGCTACGTAGTCC
ATAAGGCOGCTGCACAACACTGCACCTCTTGTGTCCAGGTGGTGCAGGAGGCGTTGGACCAGGCCAGGAAGGGCAGGACGTG
CATCGTGTAGCCCACCGTCTGTCCACCATCCAGAAOCCGACTGCATCGCTGTGTTCCAGGGAGGAGTGGTGGTGGAAAA
GGGACGCACCAGCAGCTGATOGCCAAGAAGGGAGTGTACCACATGCTGGTCAACAAACAGATGGGTCAACACAGTGGATAAA
AGACGGTGACAACATATGCCAATATCATTTTTACATCGCAATCCAGTTTAATGGGTGTGTGTAAGACTGAGAGCCTTATGTC
ATAAACGATATGCTGAGTTTACACTTACTTTTTATGTGTTCAATTTATTTAACCATTGCTTCATAGTTCATTATCGTATCAA
TCTTCTCTTTAGCTCAGACGACTCCTCAGTCAGCTCAGGTCCCGTTCATTACCATTCTGGATTATATTATCGATCATCT
TAATTCGAGCAAACCTTTCACCTTTTCTCTGTCTTCTTTTTCTGAGGAGCTCTGATGATAAAGGTGTGGATAAAGTGTCTCA
AGGGCTTTGGGTGACAGATGTTGCTTCGCTTTTCTCAGTATTTGGTAAAGAAACAATGAACTTGGCAGGTCATTTGTTAAA
TTTGAAGCAGCAATTACAAGTCACACCCCGGCGTACGTGATTTCCCTGGCAGATCATTTTTGGGAGGCAAGTTGACCAATAA
AATTTGCTCATTATGCAAAGATTCACCACGTGCTTTTTATCAATAACATTTAAATAGCTGAGGTCCATAAACCTTCATTA
TGCCATTACAGACATGTAGTGCAGAGAAGGCCCTTAGAACCGTCAATTTACATCTAATCTACTAGAACTCATGGGTGAGTGTCC
CAAACAGTGAAGGCGTTTTCTGTCATGACCGTTGGTCTCCCTAAGCTCACATACTAAGATGCCAGTGTAAACACATGTA
ACAACACAGATCTTTGATAAAAATAAAGAGAAAAACACTAGAAATGTGTTCCGTGAGGTATCATCATGGTTCGGGTCTT
CAGCAGTCAAGTCTGAGGAGGAAAACAGTCCATCAAAGTAAATGCGATTGAAATCAGCGGTGCCTCTGAAGTCAACCTAAC
GTTGAAAAAGGAATCCATAATCCCAATGTCAGAGGCAGCTGTCAAGCTTCATGCAGCCAAGGCAGCAGAACAACCTGAAG
CAACAAATAAAATAAGGTGCGTTGACGATAGAACGAGCGTTTGGGTGATACTCTAATCCAGACTGGAGCGTTCAGGAATGA
TACATTGGATTAACCTTAACCTGATGTTATAGAACGTTAGAGTGAACAGGAATCTGGGTCCAGCAGGGTTTTTAGGAACAA
CCTTTGCAACAGTCAGATCTGGCTGCATGTGCACGGTTGTGCTTCAGATAACAGGGTCCCTAAAGGTTGAGGGCCGAAGAG
ATGTTTGAATGAAGTAGAATCAGACTAAGTAAATAAAGCTACAAACTACTTTATTAGATTACAGAGCAGGTAATCACCAC
AGGACTGACAAGTCAAAGACAAAATAAGTGAGCGTTTCCCTTCTTAAGATTATTAGTCTTCTGGGACTCATCCCTACAACCAC
TACTGCAGGGATGACGCTCTCTGAAATGCTCTCATTTTTCTATCTACGCAGCCTTTAGTGTTCCTGTTTTATAATTTGTG
AGGGTCATTAATAATGATCCACAGGAACATCTTTGAAGCGCTTCTGTCTTAAGATCCCTGGGATTTTTACTTGGAAACGCTA
CTTTTTAAAAATAAATCTAAACTCTTAGATGCCAGCAGGCACTCTGAAAGGACCAAAATATGGACATGAAGAGCGAAATA
GTGAAAAACAGAATTTGAGGTAGGCTTTTTCCCTCGATTATATATTGATGTGTTGCACCTTAGAAAAATAAGAGGAATTA
AACATTCAAACCAAGAAATGTTATATTAAGCTGTTATTGCAATTCATTTTTAATTTTTTATATGTATATATGTACACAGT
TATGGCTCTGTAGTGACAACAAAATAATGTAGAGGAAGAAGAAAAGCATCAAAGCAGCGCTTATTGGTCCCATCACAT
TGGTAAGATCAGCCCTGCTCTGTGATAATTGATAAATAAGTTTTTAGCTTGACTTATTGTAGTTGTGATGCCGTTCTCTG
CCACAAATCATTGCAATGATGAATTGCATCAATTTTGATCATATTTGAGTTTCTAACCTTATTTTGCTAATGAAAATTAT
TACTTCCACATAGCCCCGGGAGTAGCATTTAACAACAGTTCTCTCCACCTCTGTGCACACGTGCACGTGTAGTTCCGGTTT
GCAGATGGTTGGGACATCTTGATGGTACCATCCGAGTGTGATGGCCATTGTGAACGGGCTGGTGAATCCTCTGATGTGTA
TOGTGTTGGTGAGTACTGACAGCTTCATCCAGGAAGCCAAACTGTCCAAAACCACAACACAAGCAACCCAGTGAGTG
TTCTCAGCTGAGGAGCATCAAGTGAGACACACAGTAAAAAAAATATTAACCTGTGTTTTACTCTAGGAGCAAACAGCAC
CTTAGAAGCAGATATGCAGAGGTACGTTCTGAGCACAAGCTTCAACAATGTCTAATGTTTTAAGAGTGTGCTGCATCATTC
GCCATCTGTAGATAATGACAGGGGAGCAATTGTTAAATCAGGGGCAGATATTTTTCAATAAAACCTCACAGGTATTTGACGC

CTGAGGAGCCAATGTCATGGTTCATTTCTTGTACTTTATAACCCCTTTTACTGGAAACCATAGAGGTCTCGCTCAGGAAAC
GTTACATTCAGACTAGGAATGCAACATGTTCCAATAACATGGCCTAAACCTGGTGTCTGGCCCCATCATAGTGGTTCTAGA
TATATTCAGCTTGGCCTGTGTGGTTCTTCAGTCTGGACTGATGTCCAGGAAAGAACTATACTATAAACTATACTATAG
TCCACCTATGAACCATAATTTAGGTCATGTACCATAATCCTCATCCTCTCCTCACACATCACCAGCCTCTCAGCAGGTCTCCA
ACCATATCGGCTCCTGTGGTTCGACATCTTGATTGCTAGTGGATAAAGGCTCGTCCCTCCTTACCTAGCCGAGTTTGGT
TACTCTAATCTCCAACCTCTATTTTTAGATTCTCCATCTACTACTCCATCTTGGGGTTTGTGTGCTGGTAGTGGGTACC
TGCAGATGTCTGTGTGACCCCTAACGGCCGGCGGGCAGGCCAAAAGAAATTCGGAGTTGTTTTCCAGCCATCATGCAGCA
GGACATCAGCTGGTATGACGTGACTGAGACAGGAGAGCTCAACAAGCGTCTCACAGAGTGGGTGACGCACATCATACACACT
CCAGTTCCTGTACAGCTGGCGTGGTGTATCATATGTGGTGTTCGATTCCCTGGTGGCAGATGTCTACAAGATCCAGG
AGGGCATCGGTGACAAGGCGGGTCTGTGATCCAGGGCCCTCCACCTTTATCACTTCTTTGTTATTGGTTTTGTACATGG
ATGGAAGCTCACCCCTGGTCACTCCTGGCCATCAGCCCTGTGTTGGGTCTCTCAGTCCCTTTACAGTAAGGTAATATCTCC
ATACACCAGTGGTGCCATAATCCAGTCTCGAGGGCCACAGCCAGCCAAAGGAAGCGTTTTCTGCCTGGTGGGACAGAAAC
CCCGGCTGGGACACGAGGGCTGGATTTGGGCACCCCCATCCATAGACAGATTACTCCCTCTTATTATACTCCTATACC
TATATTATATCTCAACCTGTGGCCCCATAGTTGCTGACAAGCTTCACCAGTAAGAGCAGACAGCGTAAGCCAAAGCTGG
AGCTGTGGCAGCGGAGGTGCTATCCTCCATCAGGACTGTGTTTGGCTTCAGTGGCCAAAGAAAAGCCATCAAAGGTGAGTG
TTCTAATGATTATTACCTTAGATTATCTTACATTTATTGGGACTGAGTCAGAGTGGGAGCAGTTAATATCCAGACATCG
CTAACTTTAGTTTTCTTTGTGTTCTCAGATATCATAAGAACTGGAGGATGGGAGGACATGGGAATAAGAAGGGAGTT
GCTGCTAACACGGCCACAGGCTTCTCCTTTCTGATGATCTACCTGTCTATGCTCTGGCCTTCTGGTACGGGACTACTCTGG
TCTCAACAAAGAGTACACCATTGGAATTTACTGACTGTGAGTGGAGGCAGACATCATCACTTCCACCATTAAGAATAAGA
GGTTGCTGCAGAAACAGTGACCAGTGTGTCCAAATGAAGGTGTTCTTGTGCTCTACGGGGCATACTATTGGACA
GGCCTCTCCAAAGTCCAGTCTTTGCCAGTCCAGAGGAGCGGGTATAAAGTCTACAACATTATCGACCAGTGGTCCA
GTGGTGTGAATGGCGTACCCAAAAGCGTACCCAAAGAAATTCGTTAAAGCAAACACGCACGCATGTATAGGGACAGCTAGC
TGTGCTCATTCTAACTCTGAAAAAACAACAAGCAAACAACAACCACCAGCTGGTTTGAAGGGCCATTTTGGTATTTTA
ATTTCAGGCTGGGCTGTTTACAGTTACTGTTAGTACCTGTTGGTATTTTACAAATCAACATATTTAAATGACTTTAGCGA
AGCGAATGCTATCACTTTGTACATACTACTGATTGATCTGTTTGCAGTGAACACAGTATGCACACAGCTTAGAGCCAACTG
GAAAAAGGCATTAATCTCCAAAGGTTCTTTGGGTCAACATGCTTAAAGTACAGATTTTAGAGGGTAACTCCTCTGC
TTTTCTCCTCATTCTGTAATAATTTGTGTTAAGGTCTTAAAAATGAAATCGCTTCTACTTTTCTACTTCTGTTCCACTCT
TGCTTTCATAGTTTCTGAGTCTCTTTTTTTTTGGTGTTTTCTCTTTCCAGGAATTGCTTTGTAGTTCCTGTTGTCTGT
TCTGTTGTGCTTGTGCTGGATTGGTTGTTAGTTCAAGAGCTTCCAGCCGTCGGTCCAACATTCACTTATAAAATAAAAT
AAAAATAATTCCTAATGGGAAACAAAGCTGGCTGACACACATTCAGATGCCAGTGTAGCCTCATGTAGCCTCACTGCTGG
ACAGTACTAGTACTGCTGGATAGCAATAGTTGATATAGAATAAAAAAGTTTTACAAAGAAAAACAATGAAGCCTTCGAC
TTCTAAAGCATTTCTGCCATATCTCAACAGAAACGAATATTGACAGCTTTTCAGAGGAGGATACAAGCCTGAATACATC
AAAGGTGACATTGTATTCCAGAACATCCACTTCAGCTACCTTCGAGGCCAGAAATTAAGTATGTTACCATCCATCACAG
TTTTCTATATGAAACAATGCATTCAGCCATGTCCCTGCCCGACTCACCTCTCTCATCCGAGATCTTAAAGCAGATGTGG
TTTCATGTGAGGAACGACAGACCAATTGCTTTGGTGGGAGCAGCGCTGTGGTAAAAGTACCACCATCCAGCTGCTGCAGA

GGTTCTACGACCCCCAGAAAGGATCCGTAAGTAATTTGGTCTATTTTGAAGTATGTTTTGAAGATTTTCTCTGAGCATTGG
CAGGGTATTGAATTTATCCAGGTGTGAGAAAGGACATGATGTCCTTCAACCATGACAGTAGGTGCACATGTGCACTATAT
AGAGAGAGGTGAAGCTTGAGTGTGCTGGCAGACAACCCGTAGTCAACCAAACTTCATTCAGATTTTTTAAAAACATCAA
ATGTGCTTTTTGGCCTCGATGAAACTAAGAGGGAAAAATGCTTAGAATCAAACGCTGCCTAGTGAATGCACTTATTTTAAAC
ATGTGAAAACACACGTGTGTAGATATTTATCGACGGTCAAGACATCCGCTCCCTCAACATCCGCTACCTGAGAGAAATGATC
GGAGTGGTCAGCCAGGAGCCCGTCTTTTTCCGCCACCACCATCACCGAGAACATCAGATAOCCGCGACTGGACGTGACGCAAG
AGGAGATCGAACGAGCCACTAAAGAGTCCAACGCTTATGACTTCATCATGAACCTCCAGACGTGAGGCCTCATTTATGGCT
TCCATATTTATCACTGGCTCCGCTCCGCTCTAAAGCTAATATTTACATTATGATCTCTGAGCAGAAGTTTGAGACGCTGGT
GGAGATCGAGGGACTCAGCTGAGCGGAGGACAGAAGCAGAGGATCGCCATCGCTCGAGCTCTGGTCCGCAACCCATAAATCC
TCTGCTGGACGAGGCCAGTCTGCCCTCGATGCTGAGAGCGAGACCATCGTACAGGCTGCTCTGGACAAGGTACAGACCCA
GACACGTGCAGAGATGATGGAAACAAAGAAACGCTGCTCCACTGCTTATTTGGTTTATGTGTACCTTTCATGGTGTCTGGA
AAAGACTCCCAGCCACAGAAGCCCTCCAGAGCCCCAGAGCTCCAAATGCCCCCTACTAAAGCATGCAGCAGGATCAAAT
AGAATCATGCTACTTTTTTCTTAACATGCTCTCCATAACTGACTGATTCTCCCGTGAGGGCATTTGTTTTAAGGTGTTTT
TAGTGCAGCTTATTTTCAGGTCTCCAGCTCTAATTCATCCAGGTCTGCTTGAGTTCTGTACTTTTGTTOCCATTCATTTA
AAAAAAAAAAAAACCATTTAAATGGGTTTTCTCTACAGGTCCGACTGGGTCCACCACCATCGTCATCGCTCACCAGCTCTC
GACCATCAGAAAACGCGACATCATTTGCTGGATTCAGTAAITGGTGAATCGTTGAGCAGGGGACTCACAGCCAGCTGATGGAG
ATAAAGGGAGTCTATCATGGCCTGGTGACCATGCAGGTATCCAAAACCTCAGAGGAGTGAAGTGGTTTCTGTCTGACTCTG
AGTGTAGTCAGAACGTCCATCCTGGGACATCTTTTCTGCTACAGAGCTTTCAGAAGCTGGAGGATCTGGAAGACTCAGACT
ACGAGCCCTGGGTCCGCTGAGAAGAGCCAGCTGATCGAATCCTTCTCCAGTCTCCCTGCAGAGGAGGAGGTCCACTAGAGG
CTCCTTGCTTGCTGTCTCAGAAGGAACAAAAGAGGAGAAAGAAAAATTTGAGTGGATCAGGACAACATAGAGGAGGTGAGC
TGTGGTCCAGAGACTTTACAAAACAGGCGGCTCGCCTCAGATTCAGACTTAAACAGTGAATAAGCCAGGATCAGCTGACCTT
TATAGGTCTGGTTATAGTTGCTGGTCCTTAACTTTCCATACATTTAAGCAGCGTTTTGTAGTGCATGTGTTGTTACTGGTA
CTCTTTCCAGTGTGTAACACAGCTTCCAATAATCTTTCTGCAGGATGAGAAAGTTCTCCCGTGTGTTCTTTAAAGTAT
GGTTACAACGTTTCTGAGTGGCGTATATTTGGTAGGAACCATCTGCGCCATGATCAACGGTGGATGCAGCCAGTGTTC
AGCATCATCTTCCAGAGATCATTATGGTGCATTACCAGTTCCTTTTACTTTCTTTGCTGTTTTTCAGCGTTTGTGACGTA
AGAGTGACTTTCTGTGAATTTCAAGTCTTCAGAGAAAAGGATAAAGAAATFATCCGGGAGAAGTGTCAATCTTTTGTATAC
TGTTTGCTGTGATGGGGTGGTACTTTTCTCACCATGTTTCTGCAGGTTTGTATCCACTGAATCCCAACGACCTCTTTGAT
GTGTGCTGTTTTACATTTTGAAGGAAAAAGCATAGAAAAAGCCCGGGTTGTCCAACAATAGCAGCGTTTAGTCTCAAT
GCAAAGGTCAGATGTAGTTTTGGTAAACACATGGATTCTATGCTATTGTCAGTCTGTCAGCACATTTCTAGAGGCTTCACT
CACCTAGTCTTTCAACACCACCCTAAGACTACACACTTTGTCACAAATGTTGGTTGTGTAGCACCTTACTAGGTGTTGG
AAATGTTCCCTGGAGATTTTGGTCCATTTTGACATGATAGCATCAGCATCAGCTGCACATCCATGATGTGAGTCTCCCTTC
CAGCAGATCCAAAGGTGCTATCGGATTAAGATCTGGTACTGTGGAGGCGTTAGAGTGCAGTCACTTATTGTATGTTT
CAGAAGATATGAGGTCTTTGATGGAGCATTATTCTGCTGGACGTAGCCATCGCTGTGGTCATAAAGGGGTGAAATGATCATT
AOCGACACTCAAGTAGGCTGCAGCATTCAAATGATGCTCAGTTGGTACTAAAGGGCTCTTTCTTTTCAACTGGATTCTTGT
GTGCTGGTTTTGGTCCAGTCTCCATTTGTATCTTCACTGTGTTTATGTTGGACAAACACTAATTAATTAAGTGGAAAGT

AGTTTTATGTAAATGACATTTGATCCTGGCCGGCAGCAACCGTCTTACAAGCCTCTAAACTTGTCTTOCCACTAAGGCTTGA
AGTGAGGGTAAACAGTAGGTCTGGAGATTCTGGGGGAAAACTGCCTGAACCTTTAATTGAGGTTAAAAGTTACCTOCAG
CCTTCCAAGGGACTTTTACCTGCTACCCCTCTACCCAGATGGCAGCATGTAAAAATAAAGCAGCCCAACATTTAAAATGGC
AATTTAATGACTAACTTTACTGATTTAACCACCTCCTTCATCAGAGAGAATAAGATCGCAGCATTTAAATTGCTTTTTATTT
TTTTAATTA AAACTTTATAAGCAGTACTTACACTTAGACTACTATAAATGATCATTTATGGATAATGATATTATGACACAT
CACATCTCTTCTAAACAGTTTTGGGGTTTCCAGGGTTTCTGTTTCAGTAAATCTGGAGAAATCTGACCCCTGAACCTCAGAC
TCAAAGCCTCATATCTATGATGAGACAGGTGAGTGAGATAAACTTAAACCTTGAAGTTGGGATTCTTTTGAACAGGA
GCATTTGGATATAATTTGTAATATGTATCTCGGGGCACTGGACTTTAAATAGATAATGTCAAATCTTGGATGAATAATAAGG
TGCCGACCCAAGACACGGCTTTTATTCCACACATGCAGATCATGCCGGATCAGCAACCTAGCCTAATTTAAAACAGTTT
TGGATTCATGTTTCATCAAATATTGATGTAAAATCTCAATCCTCTCTGTATTGAAGGACCTCAGCTGGTACGACAATCCCAA
AAACACCGTTGGCGCTCTCACCACTAGGCTGGCTGCCGACGCCGACCGTACAAGGAGTAAGTCTGATCAGGAGCTGACA
TTTGACCTCAGCGAGGCTAATTAACGGAGGATGCCCTCTCTGCTCAGGCTGCAGGGTGGCGCTGGCTGTAATGACGCAG
AACTTGGCCAACTGGGCACCAGCATCATCAGCTTCTGTACGGCTGGGAGCTGACCTGCTCATCTGGCGGTGGTGC
CCATCTGGCTGTGGCCGAGCGCTGAGGTCAAGCTGTGACAGGACAGCCCGAAGACAAGAAGGAGCTGGAGATGGC
CGAAAGGTTTTCTTTTTCAATTCCTTATGTTCCCTTCAATTTTGGGGATGTGAGATGAGCTTATTTGTTTACCCCGAAT
ATTTTCTGCGTGGTTTTCTCCCTGCAGATGCCACAGAGGCCATCGAGAATGTGAGAAGTGGTGTCCCTCACCAGAGAACC
GACATTTGTGGCTTTATAAGGAAAATCTAACTGTCCATACAAGTAGGACCAATGACACCATTCCTCATATGTGACAGT
ATTTTCATGCATTTTATGTATCCAGTGTITCTCATCACTTTTTTCTCACTCTGCGCAGGAACCTCCAGAAAAAGGCCAAA
ATTTATGGCTTAACCTACTCCTTCTCACAGGCCATGATCTTCTTTGTTAAGCTGCCTGTTCCGCTTTGGAGCTGGCTGA
TCGAAGCAGGACGGATGGATGTGGAGGGAGTGTCTCTGTAGTAOGGACACAGCACGGCCATACTGTAAAGACTCAATGCT
TTTTATTTCCAAAGGACAGCAGAGCAGACATCCCAAGTTACAACACCAACAGTAAAGACATTTGAAAGAGGAGCTTCATCTTC
TAAAGTCTTCTAAAGTTATGTTAGCCAAGCTGACATCAATTATGGACACCCCTACATGAGACTGTGGGGGCTTAAAGCAG
CTCCTCCTGCACAACATCCACAACCTTAAATGTAGCAACCAATAACCCATGGTTGGCATATTTTTCTGCACTACTTAATCATA
CTACCTCCGAAATCTCTTATTTGCACATTAATTTTCTTCATATAGTCTAATATTCCTTCTGTATATAATTTCCCTCTGT
ATATAATGTATAATGTACATAGTAATGTACATAACAAAATAGTTTTTTCTTTTAGTGTGTATATTACTGGCTACTTCTGT
AACAGTGAATTTCCCAATGTTGAAATCAGTAAAGTTTTATCTTATCTTAGTCTCTTGTAGTACAGATTGAAGCGTGTAGC
GGTGGGTGAGATGACTTTCCAGCATGCCTTGGCACAGCTGTTTTTCTAAAAGTTAACTTAATTTCTATTTAAAATCACC
ACAAATCTTAATGTTGTTGTGCACACCGTAAGTGAAGTTGCATGTTGTCTGGTTAATGAGTGGCTTGCATCCAGAAGTCAAG
TTTGTAGTTAATTAATGCTGTATAGTGTTCAGTGACGAGTTACCTGTCTATGTTCTATTTTAGCGGGATTGTGTAATAA
TTTGGAGCTAAGATTAATTCAGTTGAGAAAACGCAAGTCAGATCATGGATGGATATAATAAAAACTGTCAATTAACCTTA
TTTCCAGTGTGGTTATGACAAATGCTGTACGGCGCCATGGCTGTGGGAGGCCAACACTTATGCTCCCAACTTGGCCAAAG
CCAAAATCTCAGCTCCACCTGACGATGCTAATAAACAGACAGCCGCCATAGATAATCTGTGACAGGAGGAAGGAGACT
GGTAAATGACACAAATCCCCACACAACAGTGTGCTTTAAAATGACAGCAAGCATGCTAACACACGGTTCTTCTCTGTGCT
CTCCAGGAGAAATACGACGGCAAGTCTTTTTGAGGACGTCAGTTAACTACCGTGGGGCCGATGTGCTGTACTAC
AAGGGCTGAATCTGGAGGTGCAAAAGGGAGAACTCTGGCTTGGTGGGCAGCAGCGTTGTGAAAGAGCACCACTCCA

ATGTGCATTGTCTTTGGAGAAATGACGGACAGTTTTATATAAGCTGACATGGCCCAACACAACGCAAGTGGCTGGAGTGAGT
TTTTGCTAGAGAAAATGATCAACACCAATCCTGATTAAGCACCAACAACGCCTTTTGTCTTCAGATTCTACTACTACTATTC
TGAACAGCACGTTACAGGAGGACATGCAAAGGTGTGTTTACGGCTGCTTTCTTACTGTAACTTAAAAGCCCTCCTGAAGG
COCTGACATTGCTTATTTTAGACTTTGAACTGGCTTCTATAAGATCTTCACTCCTGTGATTGTTTGTACCTTTGACACA
AGTTCTGCTTATTTTATCTATCCAGGCATACAGTTCTGCCAGTTTCTGTGTTAACAGGTCTTTGGTGGTCTCACCATGTT
GCGTTTTTTGTCAGATTGCCATTTAATTACTCCGTCTTGGGATTTGTTGTGCTGCTGGCCGCTACATGCAGGTGTCTTCT
GGACCATAACAGCCGGGCGCCAGGTGAAAACGCATCCGCAGCTTGTTTTTCCACTGCATCATGCAGCAGGAGATCAGCTGGTT
TGAAGTAAAGACACAGGGGAGCTCAACACTCGACTGACGGAGTGGTATCAGAGAGTCCAGCTTCAGCGTTCACGCTCT
GTACGGCTACGCTCGGAGGTGTAGATGATCTGATGGACGTGCTTCTTTTTTCCAATGGCAGCGAGCTCTACAAGATCCAGGA
GGGCATCGGTGACAAGGTGGGGCTGCTGATCCAGGGTACACCACTTTCATCAGGCCCTTCATCAGCGCTTCACCAAGGGC
TGGAACTGACGCTGGTCACTCGCCGCTGAGCCCGCGCTGGCCATTTCCGGCCCTTCTTTCAGTAAAGTAAGAGGCTCCT
GCTTAAACCCCTCTGCTCTCTTTGTACACGGTTCCCTAAACACTGGCTGTGACCTCTTTTGGGAAGGTGCTTGGCTCCTTCA
CCAGTAAGGAGCAGACGGGTAAGCCAAAGCCGAGCCGTGGGGAGGAAGTGTGTCGGCCATCAGGACCGTGTTCGCCTT
CAGTGGTCAGACCAGAGAGATTGAGAGGTGCTTACATAAACGTCTTTGACCGTAAAGCTACATTTATATAAGAGCAGAAGC
ATCACATTTTCAGCTTAAAAGCTTTTTTGTGTTTTCTGATTTTATTAGTAATGTGGTTTACTGAAAGACGGTATTGCAGGG
GGTGAATATTGAGAAGCGAAGCGATTAGTTTCCATGGAGAAAAGCTGTAACCTGGATGAAGAAAACCGACAATTCAATTTGTG
AAGGACATGACATCCATCCAGGGACAGTGTATCTGCTAGCTTTGCTCTCCACTCACACCCCAACCGTTCTGGGTGTGTGCT
TCAGATAACCAGAAGAACTGGGGACGCAAAGGACGTGGGAGTGAAGAAGGCCATCTCTCCAACATGGCCATGGGCTTCAC
CTTCTGATGATCTACCTGTCTATGCTCTGGCTTCTGGTACGGGAGTACGCTCATCTGAATTTTGTAGTACACCATGGC
AATTTACTGACTGTGAGTACCTGTGGCAGCACTTTCGCATAAATGAGGGTGGTGGGGTCTGACAAGCCTGGGTGCACTT
GTGTGCTTTCATTAAGGTGTTTTTTGTGCTGCTTATTGGAGCGTTCAGCGTGGACAGACCTCTCCGAACATCCAGAATTTT
GCCAGCCCGGAGGAGCCGCTATAAAGTGTACAGCATCATGATAACGTAAGGAGACCACTGGATCCGGGACTCGTTATTA
TTCTCTGAAATTAGCATCTTTGCTGCTGGAAGCTTGTGTTCTGTGTTAGTACTGGCGGTGTCAGCACCTGGGCTCTG
GGAGTCCGCTTTGATGCAGAGAGGTGGGGGTTTTGCCATAATGGGGCACCTACTGTTTGGGTTCAATTCACAGGTCACTTG
GCTGATGTCACACAGCTGGAGCCGACTGATGTTTTCAGCTTCATGGCTCTAGCGTGGACTGGCTCATGTTTAACTCCAGT
GCTCAACAGGTGAATAGTAATGAACGAAATGCAGAAGAGATGAACCGAGAAAGTGTGACGTGCTGAGGGCAGCATTCTAGC
AGCAGCAGCAGCGCAGTGAAGTCTAATCATTAGTCTTGCACCCCGCTGCTGTTGTGAAACCTGTTGTGAOCTCACTGT
CAGCCGACAGGAAGTGTAGGTCTCTACGTCAAACCCACTTTTCCAACCACATTGACCTTTTACTTTGTAAGGAAGAGGATT
TGATTAOCATGAAAACCCACTTCTAGCTCAGCGCAGGATGTGGAGGAATCAAGTGGCTTAAATAGTCTGTCTGACTGGA
TCTGGACGAGAGGTCTGGCCCTGTGTTTACGCCTGTGTGTAACATTACCGCACTGTTTTATTTACATTTTGTGTTTCT
CCAGCTTTCTAAGCTTGGCGTCACTTTGATGGTTTGCCTTGAATCTGCAGTTTCTTTGGAAGGGGATAAATATCACCTCA
AACATGAAACAAATTGCTAACAGCCACAGGAGATGGAATAACACTTTGTGTGTTGGGGGAGGCTGTGAAAACACAGAGAGCT
GTGGAATAAACATCAGCCCTGTGAAACCCACCGGCAGCAGCGGACAGCCTTCTCTCATTGTCTACTGGATGTTATAAG
AATCGCCCGCTCTCGCTTTGTTTGTGTTTCTCGGGTAAAAATATTGCACGTTCCCACTTCTGTCTCAACAGAAGCCAA
ACATTGACAGCTTTTTCAGAGGACGGTTTCAAGCCGACTTCATCAAAGGTGACATCGAGTTCAAGAACATCCACTTCAATTA

CCCTTCGAGGCCTGAAGTCAAAGTGGGTGCCAGTGTCTCCACGCAATTGTTCCACAGACTGGGTCTGATGGCCGACTGTCTC
ATCTGCAGATCTTGAACAACATGTCTCTGAGCGTGAAGAGCGGACAGACCATTGCTTTGGTGGGAGCAGCGGCTGTGGCAA
AAGTACCACCATCCAGCTGCTGCAGAGGTTCTACGACCCCGAGGAAGGAGCTGTGAGTCAGGGAGTCCATTTCAGTTCCATTC
ATTTAGGCACCTCTGATAGCTACAGGCTACAAAGTGAACGCGTGACCATAAAAAGGAGGCCTTTGCTGCTTTTCAGAACAT
TCTGGTAAATATTCAGTAAACTATTAGTGAGGTGATGCTGGATTCACTCTCCTGTGTAGGTATTTATCGACGGTCACGACAT
CCGCTCCCTCAACATCCGCTACCTGAGAGAGATGATGGAGTGGTCAGCCAGGAGCCCGTTCTTTTCGCCACCACCATCACC
GAGAACATCAGATACGGCCGACTGGACGTGACGCAAGAGGAGATCGAACGAGCCACTAAAGAGTCCAACGCTTATGACTTCA
TCATGAACCTTCAGACGTGAGGCCTCATTATGCGCTCCATATTTATCACTGGCTCCGTCOCGCTCTAACGCTAATATTTA
CATTATGATCTCTGAGCAGAAGTTTGAGACGCTGGTGGGAGATCGAGGGACTCAGCTGAGCGGAGGACAGAAGCAGAGGATC
GCCATCGCTCGAGCTCTGGTCCGCAACCCATAAAATCCTCCTGCTGGAACGAGGCCAOGTCTGCCCTCGATGCTGAGAGCGAGA
CCATCGTACAGGCTGCTCTGGACAAGGTACAGACCCAGACATGCTCCCCCTCTCCATGTAGGACCCAGCCTTTGGATCTCAA
AGATATTTCTCAGTATGGGCAGTATGTGTTTAGTGGGTGTTAAGCCCAAGCACCCCTGGAGAGCATGAATCAGAATGTGACG
TCCTGTTGACCAAAGTCCAACAGGACGTATTTATGTAAGTGTCCAGTCATAGAATGAAATGCTTTTGAAGATGTCAAAGGT
CTACACCTCTGATAATGAATACAACCATTTCAAAGGAGCAATGCTGTTTTATTACCACTGTATTATTGACCATATTATCAA
AGGTCAACATTGTACGCTAGCGTGAGGAAAATCTTTAATCGCATCCTATTCCGGTTTTGTTGGTTTTTGTAAACTCCAGG
TCCGACTGGGTCCGACCACCATCGTGGTCCGCTCACCGACTCTCGACCATCAGAAACGCGACATCATTGCTGGATTTCAGTAA
TGGCAAAATCGTGGAGCAGGGGACTCACAGCCAGCTGATGGAGATAAAGGGAGTCTATCATGGCCTGGTGACCATGCAGGTG
ACCCCCCCCCCAACGCAOCCGAGATGAAGTTAAAGTATGTTTAGTGGGAGGGAAGGCCTTAAAATGAAATTCTCTCCTTG
CTTCAGACGTTCCACAATGTGGAGGAGGAAAATACCGCCATGTGGAGTTATCTGCTGGGAGAAGAGCCCTGTGAAAAGA
CCGTCTCCAGTGGTCCATCATCAGGAGGAAGTCCACCAGAGGGTCCCTCGTTTCCGCGTCAGAAGGAACCAAAGAGGAAAA
GACAGAAGAGGTGCGGTGCTGTCAAACGACGACGACGAGAAAGCAAATATCAGACATTTAGTTAAAGTAATATTCATCCTTT
ACCTGATCACCGTCCAGGAGTCCAGACGCTGCCGCTCTTTCTCTCTTTTACACCCTCGTTTGTGTCATCGCTCCACCGATAA
AGTTGATTTGATTGTTAGATTCCTAAATAGCAGACAGGTGCTACACTCAGGTGATAACATCGATGGTTATTACACATCCTG
TGTTCAGATTACAGTCCAATTTACACCCTGAGGGTGCATCGCTCTTCATTCTGACAATTAGATTTGAAAAATGATAACAAT
AGAAGTGAATCAGATCCATCTGCCTCGATCGAATAAAAATCAAGGCTCCGTCGCTGGTTCGCGAGGATGAAGACGTTCCCG
ACGTGTCGTTCTTTAAAGTGTGCATCTGAACATCCCGAGTGGCCCTACATCCTTGTGGGGCTCATCTGCGCTACGATCAA
TGGAGCCATGCAGCCGCTCTTCGCCATCCTCTTCTCCAAGATCATCACTGTAGGTCCAACGCGCTGACCGTGGGACGACGTC
CTCAACTGGTCAAATGTTGCGGCTTCTGCAGAAGGCTTGAAAAAAAGCAGTTTTGCTTTTTTATTGCAGGTGTTTGGGA
TCCAGACCGTGATTCTGTGAGGAGGAAGAGTGAATTCATTTCTCTGATGTTTGTGTTATTGGCTGTGTGTCATTTGTCAAC
ATGTTTTTACAGGTAAGTGTGTTTTGAOCTGATTTGTAATGCTGGATAATGTTTATTGAAATCAAATGGCCGAAAAACT
TTCAACAGGGTTACTGTTTCGGTAAATCCGGAGAGATTCTGACGCTGAAGCTGAGACTCCGGGCGTTCACGGCGATGATGAG
ACAGGTGAGCTCTGACCTGAACCTGAACCTGAACCTTAOCGCTGGCTTTGCTACGGCTACGTGCCCTTCAGGCCCGCCAC
AAATACAGTTTTGATTGAGGTTTCATCAAATATTGATGTAATAATCTCAATCCTCTCTGTGTTGAAGGAOCTCAGCTGGTACG
ACAATCCCCAAAACACCGTTGGCGCTCTCACCCTAGGCTGGCTGCCGACGCGCCCAAGTACAAGGAGTAAGTCTGATCAG
GGAGCTGACATTTGACCTCAGCGAGGCCTAATTAACGGAGGATGCTCCTCTCTGCTCAGGCTGCAGGGGTGCGCTGGCGAC

AATAATGCAGAACTTCGCCAACCTGGGCACCAGCATCATCATOGCCTTTGTTTACGGCTGGGAGCTGACCCTGCTCATCCTG
GCCGTGGTGCCCTCATOGCGGCGCGGAGCCGCTGAGATCAAGCTGCTCGCGGGTCAOCCGCCAAAGACAAGAAGGAGC
TGGAGAAGGCCGAAAGGTGGAGTTGAAGCGAGTTAAATTCGACCTTTTACAGGGCTGCCATTGTGGCATAAGAAAAATAA
AAGTGTGGTTTTCTCCCTGCAGATGCCACAGAGGCCATCGAGAACGTCAGAACCGTCGTGTCCCTCAGCAGAGAACCAAAAT
TTGAGTGTATATGAGGAGAATCTCAGAGTGGCGTACAAGTATGACATATGACATCCTAATATCATCCAGAGGATATTTA
CATTATCAACATTGAATTTTTCTCCTCCCTTCTCTTGATTTTGGGACAGGAACTCCAGAAAAAGGCCACGTGTACG
GCTTAACCTACTCCTTCTCCAGGCCATGATCTACTTTGCTTACGCTGCCTGTTTCCGCTTCGGAGCCTGGCTGATTGAAGC
AGGACGGATGGACGTGGAGGGAGTGTCTCTGTGAGTACAGACACAGCACGGCCTGATGATGGCTTCTGTATGTTGAAGGTGA
CATTTACACTGAGCACGAGGAGACGTGCAAAGGCACTGGCAGGGCAGTAAAGTGATAAATGAAATCCTGAACAGTTTGTGGC
TTCTAATCATTGACCAGGGTCAGGACAGCATCAOCTCAGACCTCAATGATTCAACACACACACATGTGTTTTCTGCGGGG
GAGCTTCACATTCTTTCAATTTAGAAAGGGATATAAAGTAACATAAATCATTAAACCGAGGTATTACCTCTCACACGCATA
CAGATTTCTTTTAGACTTTAGCATCTTGATATTTTACACAATTTTGATAAATTCATGACCACACTGGAAATAGCTTTTT
GTGGAGTTGGTAATAATGACGTCAACGCACCAAATCAATCTAGGTGGAAGAACACTGATTAATAAATTTGGTACAAATGAG
AAACTTTTTGAGAATTAATAAATAACTTCTGAAGTCTTCAAAAATAAGCACTGTACAAACATATTTATCACTTAGTCATT
AAAGGCTTTCCAGACGATATGGAACCGTTTATTAATTTATTATTCAAACATCCACAGGCAACCGTTGAAGAATGCCTTTGT
TTCATGTTTGTCTTAGCGTTATGCAATAGGGTTTGATAATGTTACGAAATGACACTAATGTTTATTTAACTTTGCTTCCAGG
GTGGTTTTCTGCGGTGCTGTACGGGCCATGGCCGTGGGGGAAGCTAACACCTTTGCTCGAACTACGCCAAGGCCAAAATGG
CTGCTTCTACTGATGATGCTAATAACAAGAAGCCGCCATTGATAACCTCTCAGAGGAGGGGACGTCTCCGGTAAACAC
GCAAGACCCCGCACAAAGGTGACAAAACCTGTTCCACACGCACGCTAACAGGCGGCTCTTCTGCCCCACTCCAGGAAAAAT
ACGACGGTAATGTGCATTTCCAGGGTGTAAATTTAACTACCCGTGCGGCGCCGATGTGACCATACTCCAGGGGTGAACCT
GAAGTGAAAAAGGGAGAACTCTGGCCTTGGTGGGCAGCAGCGTGTGGAAAGAGCAACCACATCCAGCTGCTGGAGAGG
TTTTATGACCCAGAGAGGGGAGAGTGGTGAGAAGACAAACACAATGTGCTGTGGTTTTTTTTCTCCCGTTATCTCCATAT
AACCCCGTGCCTCTTACCTACTGTTAGTTGATAATGATTAGCTGATGTGTTTTATTGCTGCTCTCCAGOGTTGCAGT
TATAGGACAGAGATCATCGAGATCTGTGCAGTACGGATGTTTGTCTTTACTGCGCTACAGTCTGTTTATGTTGATCAGTA
ACAAGATTCATAGTTATCCAAAATCTTTTACTCACAGAAACTATATCAGACCCACCACTGAGCTGCTTTGTGAGAACA
TTCTTGTAACGTCATTAATTTCCAGATCACTAACTCATAAAGTGGTGTGTTGACTCTCTGCAGTATTTGCATTGACAAAA
TGTAGCTGCAGGGATCCTCTTTCCGAACGTAGAAAGCTTAGAACAGGACTAAAAGGGCTAAAAGAGAAGAACTGAAACAAA
CATTTATGTAAACTCTGATTAACAGATTTATTTATATGCTCATTGGCAGGTTGAGAAAATTGATCACAAGCCAAATTTGCCGT
TACAGATGATTTTTTATGAATTTTGCAGCGTGGGCGCCCTGCTGTTTATACATATATAGACAAAGCATTAAAAACATT
CTCTAAAAAAGCATTTAGTTCAATTTTCAATTAACAAAAGGACATTTAAAGGCTTCTACACCTTCACAAAATTTTTTGA
TGCAGTATGACTGACATCACTGCTACCTCTGATCTAGTTTAGCTCCATAAAAATGTTTTGGTTTTTAAAGACACCATACAAAG
CTAAAGTTCTCCTTTTCCCTTTTCACTGACTGGACGGTGTCAACGTGAAACAGCTGAACATTCAGTGGCTGAGGTCTCAGATC
GGCATGCTCTCCAGGAGCGGTGCTGTTGACTGCTCCTGGCTGAGAACATCGCTACGGAGACAACAGTCCGCTCCGCTG
CCATGGATGAGATAGTAGCTTGTGCTAAAGCAGCCAACATCCACAGCTTCATCGAAGGGCTGCCTCAGGTAGGGCTGTGA
ATCAGGGGAAATGGTTGATTCCACATTTGATGATTOCCATGGAGCTGCCATGACCATTACACCATATACATACTGTCTC

TGAGCAGAGATACGACACTCAGGCTGGTGATAAGGGAACACAGCTGTCAGGGGGCCAGAAGCAGCGTGTGCCATAGCCCGA
GCCATCATCOGCAACCCAACTGTTGCTCCTGGACGAGGCCAOGTCTGCGCTCGACACTGAGAGTGAGAAGGTGAGACTTT
ATTCACACTAGTGCTACGTAGTCCATAAGGCCGCTGCACAACACTGCACCTCTTGTTCAGGTGGTGCAGGAGGCGTTGGA
CCAGGCCAGGAAGGGCAGGACGTGCATCGTGGTAGCCACCGTCTGTCCACCATCCAGAACGCGACTGCATCGCTGTGTT
CAGGGAGGAGTGGTGGTGGAAAAGGGGACGCCAGCAGCTGATCGCCAAGAAGGGAGGTACCACATGCTGGTCAACAAAC
AGATGGGCTATCACAACGACTGAGAGGACGCCGCTGTTTCCACGCTTTAGAAGAACATTTTAGAGTCTGAAATGTAATTTA
TGAAGTGTGCTTCCGTTTCTCTTCTCAGGCTGAGACATCGCTCTTCAAGCTTGCAACATTCGAGCACTTCTAT
AAAACCAACCTTTACTGCTGCTGACCTTTAACCTCACCTGCATTTAACTGCCAAACACAGTGCAGTGAAGTTGGCAG
AACGCTGGATTATTTATGCGCTGCGGTGCCGTTAACAGGCTGTTACTGCAATTATGTCTTGTAGTTAGTGTAGTTCAATT
CTTTTATTTGATCATATATTTGTGGTTGAAGTCTGAACTGAAACACTGAAAAATGTATCCTGTAAACCAGACTGGCAAAAGAC
TTCAGTTCAATTCAGAGTGAGGACGTGTTGACAGCATTTATGAAACTTGATTGATTTCTGATTGTTGGTTGTTTTAAAG
AGAATGTGGGGTGTAGACAGGACGCTCTTCCCCAGCGGGGGTCTTATGTGCTTTAOCGTTTGTGAATCAAACAGAAA
ATTTGAAACTAAACGTGACCAACAAAGCATOGATCCACGACAGGACCAGTAAGATGGAGCGAACGTGCCACCACACAGCAGA
AAGACACCAAGACAAACAGCTGATGGGTTGTTCAAGTTTAAAAAGTTTCAATTTCAAGTCTTTGTTAAAAATAACAGCCTTAA
TGCAGCTCTAATGTACAGTAAATACTGAAGAATTCATCAATAAAATGAAAAATAATTTCTAGCAAAATCAATAATCAGTCTAA
CATTATCACAGTCGGAGGTCTTGAGCTTCACTCTCATCGACATTTAACACATTCGGCATCGTTTCTACATAATCAAGGAGG
CACATTTACAGCCAAAAGAAAAGTGTACATGAACTTCAAGAGAAAAACAAAGGAAGGTTGTCATCGATCAGAGTTTTT
GTTGGATGTGGAGGTGCTCAGAGCTGAGACGTTGTTGCCAGGTGCAGCATCTCGTGCAGACAGCTGCTGAAGACGTTGAACA
GGGACACGGGTGGTCTGGCGGACGACTTCCAGGCCGATATGGAAATCACAATCCTGCGGGTCAAAGTGCAGAGCGGATA
TGACTGAATGTGGTGGTCAAACCGTGACGACAGCAGAGTGATCOGTCAGTAGGCCACCTGTCTCTCGGATACGGTAGAA
GAAGCCGTAGCCGTGGGAACCATCGGGCCACCGCGCCAGAACTGTGGTGTAGCCACCAGGCTGGACGACAGCACAAG
TTTCGCCACCGCCACTGCACAACAGACAGCATGGAAAAACAGCCCGTTTCTGAATAAAAGGTGAGTGAAGCTAACTAG
CCTTAAGACGGAACGCTAGCAGCGCTTTCCACGCACTACCTCTTGGCATAGAGCGGATCTAAGAAGAGTTCCGGAACAGGAC
GTCCCTCCTCTTTGGCGATGAGATACAGGCCGAGAAGGTGCCCTGTGCAAGCCTGCACAGGCGAGAGCGACGAOGTAAACAGGA
AATATTCGGACGCGTCTCCCCATTAGACTCAGACAAAAGTGAAGGGCGTGACTCACCTCGTCTTCTGGGCTCGGCCAT
CAGGTTGTTGTGTTTCTCAAAGGCCAGCTGCATGGCTTCTCTTAGCGTTGTCTGAAACAGAAAATCGCCAOCGCTCAAGC
CCCGCCCTCCTCGCCACCCACGGCAGGCTCTCAACACACTCGCAGACGGTCCGTCATGGCGTGCACCAATTTACCGCC
TCCAGGTGCAGGGCCTCATGGTCTCGTCTCGGTTGGTAGAACITGGGAGTTCATGGCTGTCTCGTAGCAACACCCCTGGCC
TGAGAGGAGCGGGGGGAGACATTCGGGTGACCCCGCCACCCCGGGTGGCTTTGCTGCGTGAATCCATAOCTCTGGTGC
AGTTTAAAGTACGCCAGCTGCATCGCCAGTTGGATGAACGTGTCGGGTGCAGCTTTTTCTGCTTGATGGGGCTTTTCCGA
AAGCGTGAAGGGTAAACAGACAACCTGCAGGTCTGCGCTGGACAACAGAGGGTGTCACTCCACCGGCCACACTGGCAA
CGCTAGCTCCACCTCACAAAAGTACTCACCGACTCAAAGTATTGTTTTTTCGACGGCGATGTGCTGCGGACTTTTTTCG
TCCACAGTAAACACCAGCTCCTCGGGAGGGCGGAGGACTCTGACTGTGTCCACACCTTAGAAATATCACAATGTTAATAGC
GCAOGCTTCAATTTAAAGAGCGCGAGAAATGGGAATGGTACTACCTTCCATTTGCCCTCGGTGCTTTGAATTCGCTGGTC
CAOGTACCAGCACATGGTACCAGCACCATGGGTGCTACGGCGGTGCTGCAAAACAAAATGACGTTCTCTAACCCAGAAA

TGTGTGTTTTGCGTGGTCTGAAOCTTCGCOGGGGCGCATCACAACCTCACATCACAGTTGGATCCAAAGTTCCATCTGAATA
GACGACTGAATTGTAGGATTTGTCCCCCAGCGGATGGTGGGATCACCCGTGAGAGACTCCCGTGTGAGCTGGTTCAGAAA
AAAGCAGGGGGCACAGTCAGACACCGTTTAGGCAGCCGGCTGGAGCATCCGTGATCTACTCAGCTCATAOCTTGGTGTAG
TTCTCTGGAGTGGAGTAAGGCTGGTCTCATCCAGACATATGGTGAACAGGCTGCTCTGGATGAGCTCCAGGATGGTCTCGT
TGTGCGGATCAATACTTATTAGATACTCTCGGGCTGCGGTGGATGGAGGCAGAAACAGGAAGTGAAGGACAAATGCTTGA
GCACGCTTCTGTAATATGAGAGTTCCCTTCAGAGCGAACAATAGAAACACACAGCAACCGCTAAAAGCTTCTGCAACATCGT
TGGTTTTATGAGCAGAACAAOCTGAGGGCCATACCTTCGCCAGCGGCTCTCTTTCAGTAGTGAGAGGGCTCACTCCCT
CCCCCTCTGGGTTCCCATCACAGCACTGTCTCAOCTAGCTCAGCTGCTGGCAAGAAAAAGAGAAACAAAACGCTGATAATA
ATAATGATATTGATCATTCTTAACAGATCTACATTAATGAGCTGCTGAAATGGAAGTAAATAGGGGTGTCAACCCCTGA
CTGCTCCTAGTGTTTTGGATTTTTATTTAACAGGAGGAAAGATATAATTAACAATAACCACAATAATGCACTAATAAGG
ACTTAGCCATCAAAATGGCCAAATCCCTTATTAGTTCATTAATTAGGGAAGAGGTCATGATTTGAGCAGTGTTCCTTTCTA
TATTTAGTCTAGTGTATTAGAAGTATCAGCCCTTCAAACATAGCTTTCAGGCAGCTGTGATGTCAGTACTTACTGTATAA
TTACTACCTGAACAGTCTCTGGGGCGTCAGTATTTGTCCATCACAGAGGCATCAAAATGTGAAGATGGTCCACGACAAAGA
ACCACCAAATGGGAAGGGCACCTCCCTCGAGCTCTGTCCAGAACGACGAGATGGAACCTGAATTGCTGTGCTATCAGCAT
ATAOCCGTGTGAAAGTCCGTTTCTCGCAOCTGTTTTAAAGTAGCTACGAATAGCGTCTTTOGTCACCCCGGTACTTTGC
AGGTGCAGTACAGCATCCTGAACCTGGTCCATATCTAAAGGTGTTTCGCCAGCTTTCGCGGGCCAACTCTCCCTGAAACA
AGGGAGAGAGGGCCGAGCGCATGAAACCCCTGGCAAAAGAGGCAAAAAAATTTCTGCTTCTACGAATGAGGAAOCCGTTTCAGAA
TGTGCTCACGTGCGGATCAGATTCAGTACTGCAGCATATGCCAAGAGCAAATACTGGCTCGCTCCAGGGCTGTTCCTCG
GCAGGAGGCCAGCAGTGTCCAGATAAGGTGCCGGCCCTGCGAAGTTTACAGTCAGCTGAGAGGGGCTGGCGCCCTCCAGAT
AAGCAGCGTCTAACACCATTGCTCCAGCTGGGAGGACAGCACAGTGTAAACATCTGTGAGGACAGAGGAGTCTGTGAAA
TGGAAAAACAAGAGTTTAGAGATCAGATACACACCAGTTCCCTCCTGGTTTCAGCTCTCTGCAGGAGCTCTGGTGCAGCTC
TTTTCAACACCCTGTTGGAAGTTCCCTCACAATGTCTCTGTGGCTTGAATTCCTTCTCGGAGCCAAAGGACGAACTGTT
GGGAGGAAAAACAAAATCAAATTTACCATCACCAGACCTGAAAAATAAATAGCTTTTATAAAGCCTGACACATCAGAAAAT
AGCCAGAAAATTTCTATTTTTTTGGAAATGAGATGTTTATTGAACCTTTTAAACAAAATCCCAATTTTTTTTTGATTTGTTTA
TGACATTTTTTTTGTATCAAATATGATACATTAGGCGATTTCCGCAACTTTTAGCTCAOACTATGTTAACTTTAGACACAA
AAACAAAGTTTGCCATAACATTTTGACAAATAGAACATGAACAACAACATTTTTCTCTTTGGTTTTTTGTCATAGCACTT
TTTAGCACAAGAGGCACACTGCAGCCTCTTTTACGGGTGGTGTGGAGTTGTTTCTGGAGAGGACAGTGGTCTGCCACACT
TGACTTTGGTTTAGAAAAAGAGAGCTCTCAGAATCTCAOCTATCAAATATGATACAAACGGCTTTAAAGGGTAAAGTGTAT
CTATAAGTGTGATACTTCAATTACTCTGGCAGTGAACAATGAAAGAGTACAAGAAAGTGTACAAACGTCAGTTTA
GTTTAAAAATATGCAGATTTAATTAACAAAATAATAACACAGAACTCAAATATCCCTACCTGCTTCCAGATACTTAGAAAGGCT
AGTTTCCAGTGTGGAGCGGGCAGGGGAGGCAGACTACTTTGGTACTGGAAGTCCGTTTCTGGAAGTATTCTGACAATTTA
TTAGCCATTTATTCTCCGCAATACAAATAAAGTTGTCACTAACACAAAGTGGAGGCTGATATCGTTCTGGTTGATTCTGA
ATTTAAGACGTTTAAAGCTCAAGAGCAGGTTATAGTAGAAGCATGTGAGGACATTGTGTTTGTCTGCTTAAATCTGAAATG
CATTGTGCTTTATGCAAATAAGATGGGAAAGAAAGTTATGCATGTTTCCCTCAGAAATCTTTCAGTTTGAACAACT
CCACACAAATGTAATGGAGCGTATTCTATGATTGCATTTAATGTGCGTGGCGTTAAATGCAACTAATCTTAGAAAGCTGTA

TTGTATAAGAACACAAAACGTTGAACTTTGACCTTTGAAACACTTGGTCTAATTCCAAGCGGCTAGATGTGCAGCCACATT
AATCATACAAAACAAAACATGACATTTTATTGGGCGAATTTATAAGATCAAAGCAATCTTTTATTGCTTAAGATCTTACT
AACCTTGCAGATTTATTGGCTTTAACTTACTCTGAGCAGCTATATCTTGCTTTAGCCAGTGATCACAATTTAATATAC
GTATCAGCACAACGAACCTATATAAAGCTGTGCTTTAAACGTGAAAGTAAAATTGAACTTCAAATAAACTTACTAGCTCA
GAAAATCGGTCTGCTGCTTCTTCAAGATGCGTGGATGCTTTATTTTTTGTACATAAAGCCGAATTATCAGGTTTCAC
CTTTACCCAGTGAACCGAAGGTTTGTGCTTCTGCTATTTTAAATAGTCCGCCCCGCTCTGCTGTGATTGGCTACCCCT
CCTGCACGGCGATTGGTCTGGTTGGTTGTCAACCTGATGTGTGCTCATCGCTTTGAGAATGTTTTACTTGTAGAGATTC
TGCTGTTTATAGAGTCACTGATCACTAACGCTGACAGCCTAATGGGAGACATTGCGTTCTGCACCTTGGCTAAATAATTAAGG
AAAGCAACCTGTGCAAAATGATTCCCTCCTCTGACTTCTAATTGGTCGAGGGCTCTTTAGCCCGCCCCACGCCCGCCAC
TTCCGGCACCAGTTTATGCATGCTGCCAGCGTTGTAATAAGATGTAATTTATGGAGATTACTTTTTCCCTAATCGTTT
GTGGAAATCTTTCATCTAAAGAGCTAGTCACAGTAGAGCAAGGGTTTATAAAAAGGAACACAGAAAGAAAACCTTTTAAAGAG
ATTTTATTTTCAAATGTAGCAGGAAGTTCACATTTTCGGGGCTCAGACTATTCTGAAACTACATTAATCTGATAAACTAC
AGTTTACAAATGTGTTAATAGAGTCCATTGTTAGCTGTTATAAATCAATGTTAATCTTCTACGTAAAAAGACAAAATTTAA
AAGAACTGAGATAATCGTTTAGTTACAGCACTTCTCAGATGCGTATTTGCTGATTACTTCATTTCCAAGTTGATGTAA
TCTGACATGGCTCATCTGGGACATTCCTGAAAGAGTGTGAGAAAAGATCTTTCAATTGAAAACACAATTGATCCATTTTA
TCCAAAACAATGAAATATTTATGCTTTTCCATTGTTTTGAGATGCCCGACTTTTGGACCTGCATAACTGGATCATAACTGT
AAATTAAGTGTAAAACATAACTGTAAACATCAGCAACTCCTCATTTAAACGTGTCAACCGTGCAGAAGCAACTCTCCTAACA
CAGATAGGCACTAAAGCAAATATGTGTTTTATGGTACCTTTCCTGGATTCCAGGCATCCACTTAGCTTTTCCCTTTAAAT
CCAGAAGTCCCTGGCTGTCTGCAGTGGTGAAGAGAATGATGATGAATGATGAGCCTTCTGGATGAGAAGCACATAAAAG
CATTGTTTTTTTCTGCCTACCCGATTGAAGTCTCCAACAATTGCTGTTTTGTAAGGCCATACAGATCCAGCAGCTCCTG
GTCTGTGGCTGCTTCTCACTTTCTTTCACATCCTCTGCAATCTTCTCAAATTTCTGCTGTAGACAGATGGGAAGCAGTGAG
GTCACATGATCTCTTATCCTCTAAACTGTGCAGGTCGATCTGAGCATGTTATCATTTCGGTGTGTGTGATCCAGACCTTCA
GTTTTAATTCAAATACACATTTAAATGCTGCATTTTGCAGCATTTTIGATCAAGCAAACAGCAGTTTTTCTACTTTTTCGGCC
AACAGCATGAATGCATTCACTACTGAAATGTGCATATTAATTAGTTAAACAGAACACTGAATGGGAAATCTGGGCATGGTAT
TCAAGTCATGGTGCAGAGTGCAGGAACTCAGAGAACAATAGGCTTTTATTCTGGACTGTGAAGCTTTTGTATCCAGGACGA
CCAAAATAATCCATTTTACAGAAGTCTGCGCTTATATAAACCATTGGGATGAATAAATCATTGAGCTCCATAGCATT
TAGATGTTTCTAAAATCTTCAAATCATGCAAAATAAGCGCATCAATTGTGTGCTGTGCAATGTAGATATTTATTATCTTC
TACATTCAAATGATATCGGAACCCCGACTTTTGTATAGTTTATAAATAAACTGCATAAAAGACTCAAGTGCCACTGCAGGAA
ACGGCTCATGATTTATTTGCAGAGCTATCTCGCAGAGTCCAAAGATGGATGACACCGTGTACCGGTACAGTAGAAAGAACGT
TCCACTGTGGCGGACAACCGCTATTCAATCTCAAGTGGACTCTCTAAACGTTTCATTTATTGACTTAAAAATAATCAGT
TGGTGATTTACCTGCTTGGACATTGTGGCAGTTGCAGCAGCAGCTCATTAGATAAATGAACAGTTTTCTCTGAAGCCCTTCG
CAGCTGTAAATACAGCACAGCTGACCGATGCAACCTGGCGCCAAAAGGGGAAGTGCAGGGAGGGAAGGCGACAGTTCATT
GGCTGTCTCACAGAACAGCACACAGAAATTCGGTGGTTTTCTTTAATGTTTATTTTATTGAGCATGATTTTAA
CAGTTGTTCTCTGATTGAGGAATTCGAGGATTTAAAGAAAGGATTGAGGAAATCGAGATTTGAAGAAAGTAAATCAACGTTGA
TATTTGACTGTGATTGGATTATATTATATAAACAAGACATCTCGACCTCTGTTCCATTCAATTAATGATTAAGGAAATG

TTAAAACAAACCTCGTGATTACACTGACTTTGGTCTATTTGTCCCTGTGGGTTATGGATAATGAGATTTGAGTTGAGAAGCT
ACCAACCATGGTGTAAAATAAAGTTTGAGACAAAACAGAGAAATGCOCTCAGAACATAGGCAAGTCAACATTAATTTGTCAA
AACTGAAACTGGAAAATATTGCAATCTTGAATTTGAGGAGGGCAATAATGTGTTTTAGGACAAGTGAATAATGATGCTGC
AACATACACATGTCCACCAGGGGGCAGTAATGAGTGTACAGAGGGCGGTGAGCTCACGTTGAACTCAAAGAAGAAAAACA
AGATGGGGGGCCTGTGGTTATCCCGTGTGTTGTATAGAAAAGTAATTATATCATTGTTGTTAACAGGTAACACGTTTTT
ACACGTAATTTTTAAGTATATTATGTTATTGTTTCAACTACTATTTGTGGGTGATATTTGCGTTTTACCCGGAACTAAA
TTCTATTTTTTCTCCAGTTTTTTTTTAAATGTTAATTGTAATTGACTACAACAACGATGCATTTGAAATACACTTTACAT
GTCTGAACTGTAAGTTTATATTATTGATTTTGGCTGGGTTGTATTGTATTGTTCTTCTTAGAGGTCTGATGGCTGCTGCA
GCAAAGTCTGCTAAAAGGAAAGCAAAAGGTATATCCACAAAAGACTGCTTCACTTCAACATTCAAGCCAAAATGGAGCC
CGCTCCCGCAAGAAGACATGCATTTTATCCTGAACACCTTAAAGGAAAACCTTTGTTTCCATCGGACTTGTGAAAAAAGAGCC
AAAGGTGTTTTCGCTTGGCGTAAAAGAAAAAGCAGGAAGCTGCTCAATCACAAGATTCAGACCTCCAGGTGAGCCAGGAT
GCTGCGAGTCAGGAACCTCCAAAACGTGGATGGACAGATGTGGCAGCTAGAAGAAAGCTGGCCATTGGAATCAACGAGGTCA
CCAAAGCTTTGGAGAGGAATGAGCTCAAACCTGCTAGTGTGTAAGTGTGTCAGCCACAACACATGATGGAGCACCTCAT
AACGCTGAGCACAACGAGAGACGTCCCTGCCTGCCAGGTGCTCGGCTCAGCCAGAGTGTGTCGGAGCCTCTGGGGCTAAAA
AGCGTCCCTCGCTTAGGATTACAGACAATGTCTCCCCAAGAGAGGGATGTGTTTCAGTAACGTGGTTGAAGCCATTTTACCCA
AAGTGCCACCACTGGATGTTCCCTGGCTCCAAGATACACCGCCAGTATAAACTGACGAAAACAGAGGCCAGAAGAGGAG
GCTCGAAACTGAGTCTGAGGAAGGGACGCTGTCTCTCCCAACTCTACAACCTCTCAAAGTGAAGAAAATAGTTCCCAAC
TCTGCGAGGAAAGGCAAAGGAAAGAAAAGGTCTAAGTAGTGGCACTCAATAAACTGCAGGGGATTTAGTTCACATTAAGT
TTTTTTTACACCCCATCAATATGGACTTTACTTGGTTGTCATTTTCCATCTGAATTGCTGTTTTTCATTTCTTTAAGGATC
TCAGTTGTTGTTCTTCAGGTAATTTCTCAAAATGTACACATTTTATTTGTTAGAGAAGGTCTGACATTGGGGAAATGATG
TCATCATAATTAAGCGGACATATGAAGACATCAGCTATTTTCTATTAACAATAAACTCCCTTCTATTGAGAAGAACTCAT
CAACTCTTTTCTTTGAAAGTGAATTTATTCAGTCAAATATCCATGCTGGTATATAGCTGATGTGTGTTCAAACAGATTGT
TATTTGCATGGGGTCTGAATGACGCTGGTACAGGTTCTTTAGCTTTTCCACAGCTTACAATGGATCTGAATAAAAAATCTCAT
TCTAOCCTCAGAGAAAATGGCGTCTTTCAAAGGTACACCCTCAACCCTCAAATCATAACAAAAATAAATAAGGCAGGTAC
CTCAGATTGAGTTTGCATAAAGGGTGTAGCTAATAATGCGCTGAAATGTTTACGGTAGCAGAACTGACATGCTCATTCAAA
CACAACCGACATCTGGAGGCACACCGTTTTCAGTGTTCATTGAGCAACAGATGTCTTCTGGAAGTCAATTGAACCTTCAAAC
AAAGAATTTGAACCTGGAAGCGCTCTTTTTGTCCGTCAATTTCCCGCAACAGATAAAGGGATAAGAGGAAATTTACATGATA
TACAGAAAAATAGCAGCGGCTCCACACAGAGCCACAAACCAACCAAAAGGGTAACTTGAACGAAAATACAATTTTCA
AACATCTGTAACCTTGTGTAGGTTGGCAGGATCAGCGCAGGTGCTCATCCTGCTCTCAGAGAAGTTATTTAAGGTCAAGA
CTTAAGACATTTTACAAATAATCTGAATTTACACATTTAGCTGCACATCACACTAAAAATGACCCAAATTTCCCCCCCC
CCCCCCCCCCAAGATGCACAAATACATTAAAAAAACTCCAGACAATAAAAAATGAGACTGGAAAAACAGGAAATGCTTAAA
TTTGAACCTGTTGCTGTAACAGAGCTGTTTATTCAGTCGCGCTGCAACATGCTGATGTCCTTCAACCTTGTGAGCTTTT
TCTGCCAAACATCTCAATAACAGGCTACTGAAAAAAAGGGACTGAAAATGGTCCATCGGAGAGAAAAGTGAAGCTGGCTGA
TTTAAACGAGACGTAACATTTACTCTTAAAAAAACACAAACCAAAAGAGACCCCGGCTACAGTGAACCGCGTGTGCTT
CAGCTGCTTTGGCTOCCAAATGAATCGAACGAGATAAATCGAACTGTACTACAAGATCTGCTGAGTGAACCAGTGTA

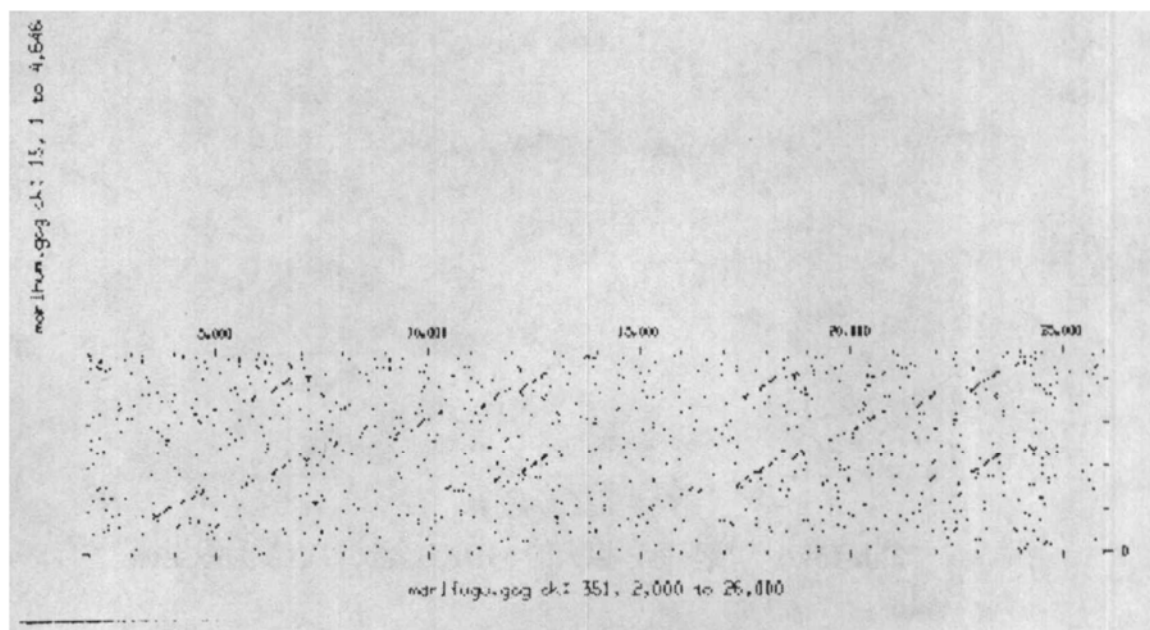
ACAAAACTGATGGTCAAATGCCAGTTACAAAATACAACCTCTACAAACAGGTTTATGGTGAAAAACAGCTCTTATTCTC
AGTAAAAACCAGTCTGCCAGTTTGAACCTCACCCCTCAATGATATATTTGGCCTGTTTAGGAACTCAAAGTTGTGCTTC
GTTTGGTAATACATTTCATCAAAACCCACGTCGAGTCTGAACAAAAGCAGCGGAATCTAATAAACTGATGGGAAAAACACTT
AAAAACAGCCCTCCTGAACTCCAAAATCACAACACTGGAAACAATAATCTGATCTTACATGTAATGGAACAAAAGTTTCAT
GTAATGCACCGGGACCAGTCTCTGCTCTTGAGTTTTGGTTTTGATGGGTACCGGGCGAATCGTAACTGTCACTTTCT
CTCTCCATCACTGGGGCTCTCTGCTTTTTGGTCACATCTGCCGTGACACTTTAGAATTCGTCTGCTGCCATTTCAATTTG
AACGGAAGCCCAATTTCTGCCAATGTGATGAGAAATAAAACGTCGCCGTCTCCGACATGTTGTAGTTTATTAATTTGGG
TCCAACTGCTGGAGAGAACCAGTTCGCCACTCTTAGCACCTTAACTGTACCAAGGCCAAGAGAGAAAAGGGACATTTGC
CTTCAGCTTTCAGTCAGAAAACTCATTTTAACTAGACTTGTGCGGTTTCAGATCGTGGCTTATCTCAACTAGTCTCT
GGTTCTTAAATACAGGCTTGACCAGAGTTCATCCTCTCTTACCGGTACATGATCTTGCCAGCAAACCGCACAAAAGT
CTTTGAAGAACAAAAGCAGCGATGTCTCTTACAAAACCTTAAAGACGTCGTGTGATGCAAATTTGTGCGGCTGCATCGAGC
ACTAATAATCTGCACCGTGTGCGGTGATCAGCTTAAACACAGTTCAGTCAATGTCAGTTATGTGAAACGGAACCGCGCG
CGCCCCCCCCCGCATGAAGAGGAGGAAGAGCTAAACAAAATTCACAAATGTCAGGCATCCTCTGTATGTGCACTGTTT
TAAATGCGTGTGTCATGTCAGGAGCGTTGACTGTACGAGTATCTCAGGCATAGCATGACTTCACATTCAGGATGTTGAT
GATTGACGAGTTGTCCCCCCCCCTCCAGCGAACTGAGGGGGGGTGGTGTGTCAGAGAGGGGGGGGGGGGGGGGACGCCA
CTTGGCCTCCGCGTGGTTGGATTCTTTTCATTTTCAGGGTTTGGGCTACCTGCATGTCCAGGTGACAACTACCACGGTCCG
TGGTGACCTTTGGCCAGTGTGTTGTGTAGCAGCCCTGAGGAACCTGCTACTGAAGGACGAGGCGGACCTGAAAGGCAACCA
CGGCTGGAGTGAGCAGAAACACCGACGTCGTGATATCGAAATACAGGCGTTAATAATCTGACCTTATCAGGCTCCATAGAT
GGACATTTCCAGTTGTAGAGGTAATACTGCAGATTTCCATCTCCTATACCAAACCTGAGCTTCTCCAGGAACACTTTATTCT
CCATGAGATCCAGGGCGTTGAAAACATCGAACCTTTCTGTGAAAACACACAGAACAGGTAAGACATGGAACCTCCAACGC
GCTCCGAGCCCCCAGACGGGATGAGATTACAGTTTGGCCAGGATCAGTCCGTCGTTTCATCAAAATCCAGGAGAGGGGTTG
TGTGTGAACGTTGTA AAAAGAGTAGGGCCCTCAGGCTCCTGTGGAGAGGGTGGTGCATCACAGTCGAGGGCAGAGTGTAG
AACTAGCAAAGTCTGTGACGCGCCCCCAGCACCCCTGAGGAGAGAGCGGCACAACGCAATTAACGGGTTAACGTACGAG
GCACTCTGGGTCTGAGTCTCAGGCTGATCCCGCTACCTCCACTACATAAGTGTCAATTAATGTTGCTCGGCAGGAAC
CAGTGAGACACCTCCTCTCCGTCATGGAAGGTGCGAGCTGGAAGCGTTTCAGGAATTTCTGTAGCAGCTCTGTGACCTGGC
GGATGTGCGGCTCTCCATTGGCCGACACCGGGAGTCTTCGTGCTCTACAAGACGCGGGGGGAGAGAAAACGCATCGA
TGGGGGTTGTGGAGCGTTGACCTGAGGTCAGGTTGGTGTGCTTAAGTCTGGTAATCTGTAGAGCTTCAATGGTCCGTTGCA
GGTTCATGTTTCTGCTCAGGTGGGAGAACTTCACTTCCACAAGCTTCTGGGGTTCAGAGAACGGTGCCAGTACCTACAGAA
AGGAGAAGAAGGTAAGATGTTGAGACTTATCCCAACACAGGGAGGGCAGCCAGTCTCTGCTGAGACCCATCCAGCG
GGCTTCTCCAGACCAGGTGACCGCTGGCACACCACTGACTTCTGTTAGAAAGGAAACCCCGGCTGGATCGGCTCTA
ACGTGTGTGTTGTGTAGCGCACCTGCAAGTGGACACGGGTTTGGGCAGAACCACTCCTGCGGTGTAACGGCCTGAAATATG
CCCTCTAGGTTCAACCTCCGCGTATCTCCCTGATGAGCACCGGGGACGCGCTTCGAAACGAGCTTCTGTGCACACACA
GGAAGTTGATTTCCACCATCCTCTTCACTCTGGAGCAGCACAAGCAAAGAAGGCTGGCAGTGAATGGCGCGGGGGCCGCA
GCAGGCAGGCAGGCGGGGGGGGGGGGGGGGGGGCTCAGTGTCTGATGCGGATGTCAGCGGAATGGCGCTGATGAAGCC
CACCAGCTTCTGTTCCAGCAGACTCCACCGCGAGTGCCTGAGGGGAGCCAGCCCGGGGACCGCAGAGCCCTGGGGGG

GACACCAGAGGTTACAGCCCCCTGOGGACGCTGGCGTTAAAGATCCAGGGAGGAAAGCAGAACCGGGAGCTCACCATTGAA
 GAAAGTTTGGCGAATAGTCGAATCTGAACATGTTGTGCTGCTCCACGTAGTTCTGTTTCAGTAAOGTGTACAACCTCTT
 CAGCTGGAGGAGACAACATTOGTTAGOGGGGCTCTCOGCTGCCCCGGGCTGCTCTTTAAAAACAGACTTCTCAGCATT
 CTGTGCTGGAOCCAGAGTCTAOGTACCACTTCTGCGCTGCCAGATCCAGOGTGTCCACATAAAAACTTGAGGTAAAGAA
 TATGGCTOCTGTCGAATGTTTTCTTTGTGGCTCTATTGGCCCCGACTCGTCACCACCTCGTCTGGAACAATGGCACAGT
 CTCAATCAAGATCTTTCCCTGACATCTGGTGCCTTTGGCAAAGATGACACTGTTAATTAAAAAACAACAACCAAAAA
 CACTAATCTACAAAGAGAGTGGGTGAGCAACGAGTTAGAAAAGAAATGGTGAGAAATTTCTGTCTCOCTCTTCTGGCAGTCAT
 GGTTATTACACAAACATTGCCACTCTGATTGCCATGACAACTGTGTCAGTACCTATGTGCACACGAATATCACTATTGATGC
 ACTTTAAATGAATTAGACACCAATTCATCGTAATCAGAAGGCACCTTACTTAACTTGGGTACAGGCTGOGTGTCCAGAACTG
 GTACTTGTGCTTGGCOGCTGTGATGCTCTTTGCAGGACCTGACAGGAGAGTAACTCCATGGCTOGTTGAATGTCTGCG
 AGCTTCTGAATGGCAAAGTAGGGTTCTAAATAAAGTCAGACATAAATAACAATGAAGGCGACGACCGTTATGAATAAAAAAT
 GTCTCAACCAACATGCCCCCGTGTAAAGTTAGACTTTATGTCGTTCCACATAGGCTGTTAATGGTCTAACGGTCTGT
 TCAGTGTGCACAACAGGTGCATTCACAATGCATTTAAAGGGATTTTCCAAATATTAAGAGGAGTTTCAGTTCAGCTGTCCA
 CTAGAATCTTTTTTGTGGGGGGGGGCGAAATATTGTTATGTGATTAATGCAATTAAGTGAATTAATCTCTAATTA
 TTAGATTAATTGGTTCAATOCAGTCTCACCATTATATACGGTATATAAGAGCAACATAAGTACTGTAACCGGAGAAGG
 TTCTATATTAACTATATAGTTGAGCTTTTGCAACCTTTAGGGTGTCTAGTACAAATCCAGATGTTTGAAATAAATCCTTAAA
 AATTAGAGAATGACTTCGTCTACTTATTCAAAACGTGATGGATTAGTACAGTAAATATTATCCCTGTTTGCAAATGTTATTT
 AGTGTCCGAAGGTTGTTCTGGCTGGAACACATTATAATAAATCATCATTTCTTACCTGAACTCCTGGGAGTCAGAGGCA
 GAGTCGGACTTGGCTCCCCCGAACTGGCTTCTCTTTCTCTCTCTGTTCTTCTCTCTCTCTTTAGCCCCAAATCCC
 CTCCCGACTCCTGCAOGTACGAGCACAACCGACGAAAATAACGGTGACTTTGGCTGACTCACAATCCAGCCTCAGTCAA
 AGCAGCGGAGAGCTAACGTTAGCTACCATAGCAACGACTAAAAGTTGAGCTAACCTAGCCTGAGCTAAATGAAOCCGAGGCT
 GTCATGTCAATCAACCAGAATGTTAAATGGCCAAATAATGATGCGCATGATGGAACGAGTTCCACCTACAATTTAGTAAAC
 GCOGCGACTGCTCCAGGCAGAATTTCTCTTATGTTTCTGGGAAAGCGACCTTATAOCCAACAGTCCAGCGATGTTTGGTC
 AGGCTAGCGTTAACCCCTGTGCTAGCTAGTCTGACCGAACACATCCAAACAGGCAGCGCCTAATATAATGTTCCCTGGAT
 GTCOGAAGGGACAAAACCAGCGCTCTCACCCTGCAGATGCTCATTCTCTCTCTGTTGTCCCGTCTATCCCGCATGTGT
 CCTGGTGTCCAGCTCCAGGCTCTGCTGGCTAGCTGGGATTCACCTGCTCCCGCATCACTAAAAATGTGTGGAAATAATCG
 OGAGTGTCAGTGGAACCGACCGGGGCGAGATGAAGAGACACAAAGCGGAAATCOGGGAAACAGCTCCGCTAGCGGCCAC
 ACGGGGATCGACCGCCAAAAAGCTCCTTTATTTTGAACCTGCAGAATGTCAGGTTTATCCCTAATTAACCCACTAAA
 TGGTTAAACGAACACTTAAAAATAGATTTGACTAACTGAGTGAAGCTCAGCCTATCGTGGCTGATAATATATTTCTACTT
 AATATTATTTCTAAATGTGTGCGAAATGTAACACCTTATTNGACATCTATGTGGACGGAGTCAACACACANCANAAAA
 CACCTCATGGTGACTTT

由于 contig1 的片段很小且与数据库中的数据无同源性，故本文不再分析。

II. 河豚鱼基因组多药耐药基因家族的阅读框架预测

一、使用 GCG 软件包（北京大学生物信息中心）的 dotplot 软件，以 contig2 为横坐标，人的 MDR1 全长 cDNA 为纵坐标，分析结果如下：



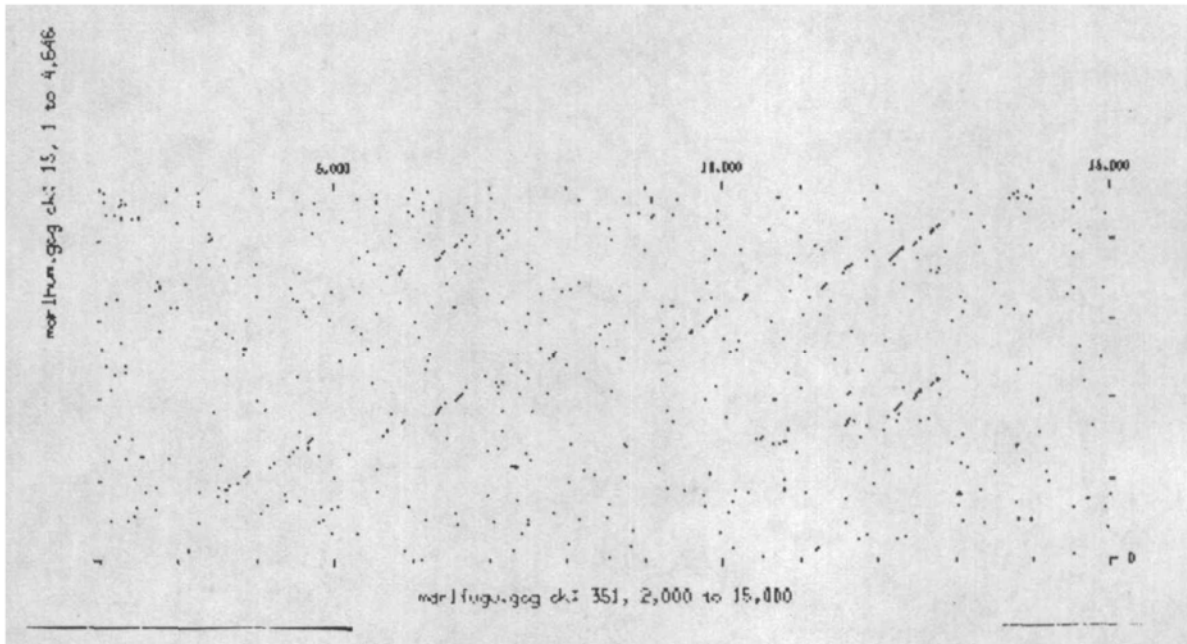
结果显示: contig2 中含有 2 段完整的序列, 都分别与人的 MDR1 全长 cDNA 有着明显的同源性趋势。

分别选择这两段序列, 即 2kb-15kb 和 12kb-26kb, 为横坐标, 以人的 MDR1 全长 cDNA 为纵坐标, dotplot 分析后显示同样的结果。

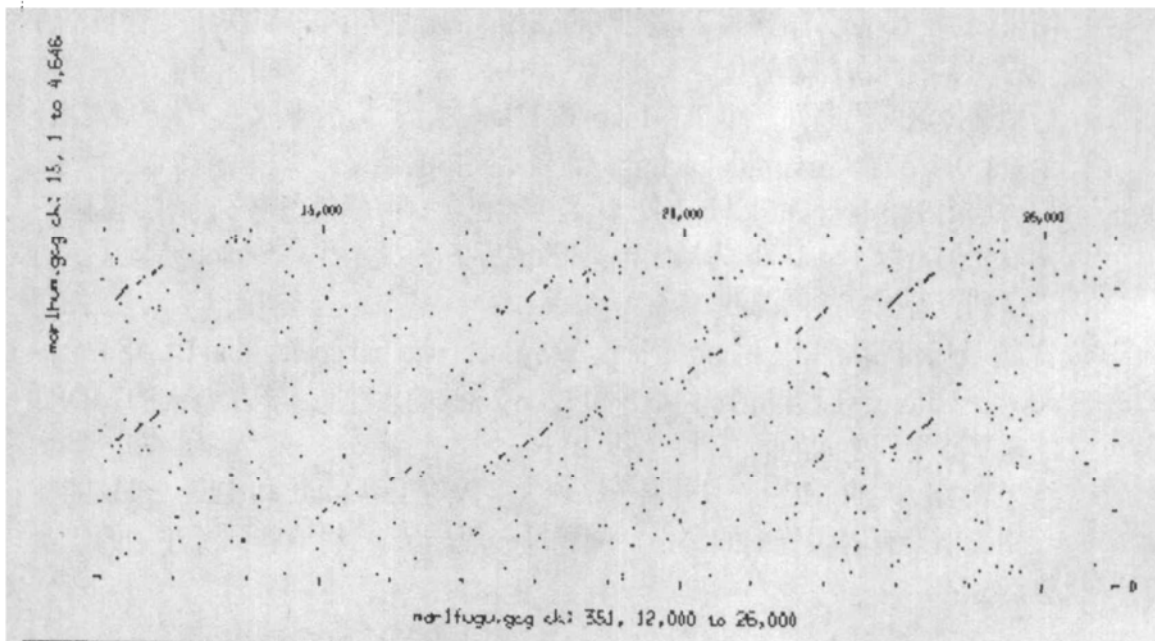
最后选定 1bp-14kb 和 12kb-26kb 作为可能包含有河豚鱼完整多药耐药基因的两段序列, 以 1bp-14kb 为横坐标, 12kb-26kb 为纵坐标, dotplot 结果显示两段序列有着很高的同源性。

二、使用 blastx 软件 (网址: <http://www.ncbi.nlm.nih.gov/cgi-bin/BLAST/nph-newblast?Jform=1>), 分析 1bp-14kb 和 12kb-26kb 的六种阅读框架与人的 MDR1 的氨基酸序列之间的同源性, 结果如下:

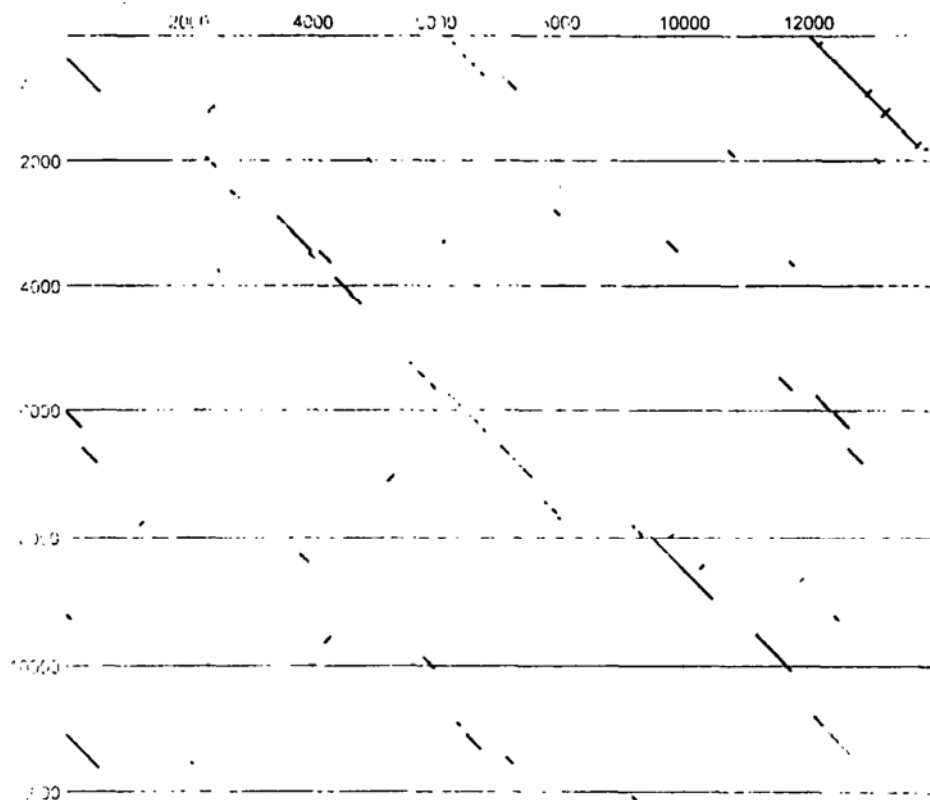
结果表明: 人的 MDR1 的外显子序列大多数与 1bp-14kb 和 12kb-26kb 的某一阅读框架有着极强的同源性, 同源性一般>50%, 将这些同源序列的起止位点标明。



2kb-15kb 为横坐标, 以人的MDR1 全长 cDNA 为纵坐标



12kb-26kb 为横坐标, 以人的MDR1 全长 cDNA 为纵坐标



1bp-14kb 为横坐标, 以 12kb-26kb 为纵坐标

1bp-14kb 的 blastx 结果

sp|P08183|MDR1_HUMAN MULTIDRUG RESISTANCE PROTEIN 1 (P-GLYCOPROTEIN 1)

Length = 1280

Score = 148 bits (371), Expect = 2e-34

Identities = 78/135 (57%), Positives = 95/135 (69%)

Query: 66 RYDTQAGDKGTQLSGGQKQRVATARAIRNPKLLLLDEATSALDTESEKVRLYSH*CYVV 245

++DT G++G QLSGGQKQR+AIARA++RNPK+LLLDEATSALDTESE V

Sbjct: 519 KFDTLVGERGAQLSGGQKQRIATARALVRNPKILLDEATSALDTESEAV----- 568

Query: 246 HKAAQHCTSCVQVQEAALDQARKGRTCIVVAHRLSTIQNADCIAVFQGGVVVEKGTHQQ 425

VQ ALD+ARKGRT IV+AHRLST++NAD IA F GV+VEKG H +

Sbjct: 569 -----VQVALDKARKGRTTIVIAHRLSTVRNADVIAGFDDGVIVEKGNHDE 614

Query: 426 LIAKKGVYHMLVTKQ 470

L+ +KG+Y LVT Q

Sbjct: 615 LMKEKGIYFKLVTMQ 629

Score = 161 bits (404), Expect = 2e-38

Identities = 85/137 (62%), Positives = 99/137 (72%)
 Query: 66 RYDTQAGDKGTQLSGGKQRVAIARAIIRNPKLLLLDEATSALDTESEKVRLYSH*CYVW 245
 +Y T+ GDKGTQLSGGKQR+AIARA++R P +LLLDEATSALDTESEKV
 Sbjct: 1164 KYSTKVGDKGTQLSGGKQRVIAIARALVRQPHILLDEATSALDTESEKV----- 1213
 Query: 246 HKAAAQHCTSCVQVQVEALDQARKGRTCIVVAHRLSTIQNADCIAVFQGGVVVEKGTHQQ 425
 VQEALD+AR+GRTCIV+AHRLSTIQNAD I VFQ G V E GTHQQ
 Sbjct: 1214 -----VQEALDKAREGRTCIVIAHRLSTIQNADLIVVFQNGRVKEHGTHQQ 1259
 Query: 426 LIAKKGVYHMLVTKQMG 476
 L+A+KG+Y +V+ Q G
 Sbjct: 1260 LLAQKGIYFSMVSQAG 1276

Score = 153 bits (383), Expect = 6e-36
 Identities = 78/161 (48%), Positives = 101/161 (62%)
 Query: 3445 YSNLQPLFFRFSIYYSILGFVAVLVAYLQMSLWTLTAARQAKRIRELFHFGIMQQDISWY 3624
 + NL+ R++ YYS +G VLV AY+Q+S W L A RQ +IR+ FFH IM+Q+I W+
 Sbjct: 104 FNNLEEDMTRYAYYYSGIGAGVLVAAYIQVSWCLAAGRQIHKIRKQFFHAIMRQEGIGWF 163
 Query: 3625 DVTETGELNTRLTEWVTHIIHTPPVPTAGVVVVICGVRFPGAHDVYKIQEGIGDKAGLLI 3804
 DV + GELNTRL+ DV KI EGIGDK G+
 Sbjct: 164 DVHDVDELNTRLTD-----DVKINEGIGDKIGMFF 194
 Query: 3805 QAASTFITSFVIGFVHGWKLTLVILAISPVLGLSAAALYSKV 3927
 Q+ +TF T F++GF GWKLTLVILAISPVLGLSAA+++K+
 Sbjct: 195 QSMATFFTGFIVGFTRGWKLTLVILAISPVLGLSAAVWAKI 235

Score = 67.5 bits (162), Expect = 5e-10
 Identities = 39/151 (25%), Positives = 72/151 (46%), Gaps = 2/151 (1%)
 Query: 3475 FSIYYSILGFVAVLVAYLQMSLWTLTAARQAKRIRELFHFGIMQQDISWYDVTE—TGEL 3648
 FS+ + LG + +LQ + KR+R + F +++QD+SW+D + TG L
 Sbjct: 755 FSLLFLALGIISFITFFLQGFTEGKAGEILTKRLRYMVFRSMLRQDVSWFDDPKNTTGAL 814
 Query: 3649 NTRLTEWVTHIIHTPPVPTAGVVVVICGVRFPGAHDVYKIQEGIGDKAGLLIQAASTFIT 3828
 TRL A+D +++ IG + ++ Q + T
 Sbjct: 815 TTRL-----ANDAAQVKAIGSRLAVITQNIANLGT 845
 Query: 3829 SFVIGFVHGWKLTLVILAISPVLGLSAAALYSKV 3927
 +I F++GW+LTL++LAI P++ ++ + K+
 Sbjct: 846 GIIISFIYGWQLTLLLAIVPIIAIAGVEMKM 878

Score = 40.2 bits (92), Expect(3) = 4e-32

Identities = 19/42 (45%), Positives = 26/42 (61%)

Query: 4133 LLTSFTSKEQTXXXXXXXXXXEVLSSIRTVFAFSGQRKAIKR 4258

+L+SFT KE EVL++IRTV AF GQ+K ++R

Sbjct: 235 ILSSFTDKELLAYAKAGAVAEVLAIRTVIAFGGQKKELER 276

Score = 86.2 bits (210), Expect(3) = 4e-32

Identities = 39/59 (66%), Positives = 50/59 (84%)

Query: 4375 RYHKNLEDARDMGIKKGVAANTATGFSFLMIYLSYALAFWYGTTLVLNKEYTIGNLLTV 4551

RY+KNLE+A+ +GIKK + AN + G +FL+IY SYALAFWYGTTLV+ EY+IG +LTV

Sbjct: 276 RYNKNLEEAKRIGIKKAITANISIGAAFLIYASYALAFWYGTTLVLSGEYSIGQVLT 334

Score = 55.8 bits (132), Expect(3) = 4e-32

Identities = 23/40 (57%), Positives = 35/40 (87%)

Query: 4628 MKVFFVVLGAYIIGQASPNVQSFASARGAAYKVYNIIDH 4747

+ VFF VL GA+ +GQASP++++FA+ARGAAY+++ IID+

Sbjct: 332 LTVFFSVLIGAFSVGQASPSIEAFANARGAAYEIFKIIDN 371

Score = 60.1 bits (143), Expect = 9e-08

Identities = 23/38 (60%), Positives = 34/38 (88%)

Query: 5608 KPNIDSFSEDGYKPEYIKGDIVFQNIHFSYPSRPEIKV 5721

KP+IDS+S+ G+KP+ IKG++ F+N+HFSYPSR E+K+

Sbjct: 372 KPSIDSYKSGHKPDNIKGNLEFRNVHFSYPSRKEVKI 409

Score = 66.3 bits (159), Expect = 1e-09

Identities = 34/69 (49%), Positives = 45/69 (64%)

Query: 5727 TIHHSFLYGNNAPSHVPAPTHLSHPQILNDMSFHVRNGQTIALVGSSGCGKSTTIQLLQR 5906

T+ + +G F++ PT P +L +S V+ GQT+ALVGSSGCGKST +QLL+R

Sbjct: 1030 TLEGNVTFGEVVFNY---PTRPDIP-VLQGLSLEVKKGQTLALVGSSGCGKSTVVQLLER 1085

Query: 5907 FYDPQKGSV 5933

FYDP G V

Sbjct: 1086 FYDPLAGKV 1094

Score = 67.9 bits (163), Expect = 4e-10

Identities = 35/62 (56%), Positives = 47/62 (75%), Gaps = 5/62 (8%)

Query: 5751 GNAFSHVPAPTHLSHP——QILNDSFHVNRNGQTIALVGSSGCGKSTTIQLLQRFYD 5915
 GN F +V H S+P +IL ++ V++GQT+ALVG+SCGKSTT+QL+QR YD
 Sbjct: 390 GNLEFRNV——HFSYPSRKEVKILKGLNLKVGSGQTVALVGNVSGCGKSTTVQLMQLYD 445
 Query: 5916 POKGSVS 5936
 P +G VS
 Sbjct: 446 PTEGMVS 452

Score = 187 bits (471), Expect = 3e-46

Identities = 99/155 (63%), Positives = 114/155 (72%)

Query: 6255 IFIDGHDIRSLNIRYLREMIGVVSQEPVLFATTITENIRYGRDVTQEEIERATKESNAY 6434
 + +DG DIR++N+R+LRE+IGVVSQEPVLFATTI ENIRYGR +VT +EIE+A KE+NAY
 Sbjct: 451 VSDGQDIRTINVRFLREIIGVVSQEPVLFATTIAENIRYGRNVTMDEIEKAVKEANAY 510
 Query: 6435 DFIMNLPDVRPHLWLPYLSLAPSRNANIYIMISEQKFETLVGDRGTQLSGGQKORIAIA 6614
 DFIM LP KF+TLVG+RG QLSGGQKORIAIA
 Sbjct: 511 DFIMKLP———HKFDTLVGERGAQLSGGQKORIAIA 542
 Query: 6615 RALVRNPKILLLDEATSALDAESETIVQAALDKVQ 6719
 RALVRNPKILLLDEATSALD ESE +VQ ALDK +
 Sbjct: 543 RALVRNPKILLLDEATSALDTESEAVVQVALDKAR 577

Score = 145 bits (361), Expect = 2e-33

Identities = 82/155 (52%), Positives = 102/155 (64%), Gaps = 2/155 (1%)

Query: 6255 IFIDGHDIRSLNIRYLREMIGVVSQEPVLFATTITENIRYGRD——VTQEEIERATKESN 6428
 + +DG +I+ LN+++LR +G+VSQEP+LF +I ENI YG V+QEEI RA KE+N
 Sbjct: 1094 VLLDGKEIKRLNVQWLRRAHLGIVSQEPILFDCSIAENIAYGDNRSRVVSQEEIVRAAKEAN 1153
 Query: 6429 AYDFIMNLPDVRPHLWLPYLSLAPSRNANIYIMISEQKFETLVGDRGTQLSGGQKORIA 6608
 + FI +LP+ K+ T VGD+GTQLSGGQKORIA
 Sbjct: 1154 IHAFIESLPN———KYSTKVGDKGTQLSGGQKORIA 1185
 Query: 6609 IARALVRNPKILLLDEATSALDAESETIVQAALDKVQ 6719
 IARALVR P ILLLDEATSALD ESE +VQ ALDK +
 Sbjct: 1186 IARALVRPHILLLDEATSALDTESEKVVQVALDKAR 1222

Score = 81.5 bits (198), Expect = 3e-14

Identities = 38/57 (66%), Positives = 47/57 (81%)

Query: 7088 QVRLGRTTIVIAHRLSTIRNADIAGFSNGEIVEQGTHSQLMEIKGVYHGLVTMQVS 7258
 + R GRTTIVIAHRLST+RNAD+IAGF +G IVE+G H +LM+ KG+Y LVTMQ +

Sbjct: 575 KARKGRTTIVIAHRLSTVRNADVIAGFDDGVIVEKGNHDELMKEKGIYFKLVMTQTA 631

Score = 69.5 bits (167), Expect = 1e-10

Identities = 32/55 (58%), Positives = 41/55 (74%)

Query: 7088 QVRLGRTTIVIAHRLSTIRNADIAGFSNGEIVEQGTHSQLMEIKGVYHGLVTMQ 7252
+ R GRT IVIAHRLSTI+NAD+I F NG + E GTH QL+ KG+Y +V++Q

Sbjct: 1220 KAREGRTCIVIAHRLSTIQNADLIVVFQNGRVKEHGTHQQLLAQKGIYFSMVSVQ 1274

Score = 74.5 bits (180), Expect(2) = 2e-16

Identities = 29/50 (58%), Positives = 43/50 (86%)

Query: 7753 DENVPPVSFFKVMRYNVSEWPYILVGTICAMINGAMQPVFSIIFTEIIMV 7902
DE++PPVSF+++M+ N++EWPY +VG CA+ING +QP F+IIF++II V

Sbjct: 689 DESIPPVSVFWRIMKLNLTWPYFVVGVFCAIINGGLQPAFAIIFSKIIGV 738

Score = 34.8 bits (78), Expect(2) = 2e-16

Identities = 14/33 (42%), Positives = 22/33 (66%)

Query: 7985 REKDKEIIREKSSFFCILFAVMGVVTFMTFLQ 8083
R D E R+ S+ F +LF +G+++F+T FLQ

Sbjct: 741 RIDDPETKRQNSNLSLLFLALGIISFITFFLQ 773

Score = 49.6 bits (116), Expect(4) = 5e-40

Identities = 21/29 (72%), Positives = 26/29 (89%)

Query: 9566 KDLSWYDNPKNVTGALTTRLAADAAHVQG 9652
+D+SW+D+PKNT GALTTRLA DAA V+G

Sbjct: 799 QDVSWFDDPKNTTGALTTRLANDAAQVKG 827

Score = 72.6 bits (175), Expect(4) = 5e-40

Identities = 36/69 (52%), Positives = 44/69 (63%)

Query: 9726 AAGVRLAVMTQNFANLGTSHIISFVYGWEXXXXXXXXXXXXXXXXXXEVKLLTGHAEDK 9905
A G RLAV+TQN ANLGT IIISF+YGW+ E+K+L+G A +DK

Sbjct: 828 AIGSRLAVITQNIANLGTGIIISFIYGWQLTLLLAIVPIIAIAGVVEMKMLSGQALKDK 887

Query: 9906 KELEMAGKV 9932

KELE AGK+

Sbjct: 888 KELEGAGKI 896

Score = 46.5 bits (108), Expect(4) = 5e-40

Identities = 22/35 (62%), Positives = 28/35 (79%)

Query: 10028 QIATEAIENVRTVVSLTREPTFVALYEENLTPYK 10132

+IATEAIEN RTVVSLT+E F +Y ++L VPY+

Sbjct: 895 KIATEAIENFRTVVSLTQEQKFEHMYAQLQVPYR 929

Score = 60.5 bits (144), Expect(4) = 5e-40

Identities = 27/48 (56%), Positives = 36/48 (74%)

Query: 10227 RNSQKKAKIYGLTYSFSQAMIFFVYAACFRFGAWLIEAGRMDVEGVFL 10370

RNS +KA I+G+T+SF+QAM++F YA CFRFGA+L+ M E V L

Sbjct: 929 RNSLRKAHIFGITFSFTQAMMYSYAGCFRFGAYLVAHKLMSFEDVLL 976

Score = 103 bits (255), Expect = 7e-21

Identities = 56/148 (37%), Positives = 86/148 (57%)

Query: 11241 SVVMTMLYGAMAVGEANTYAPNFAKAKISASHLTMLINRQPAIDNLSEEEARLVNDTNPH 11420

+V ++L GA +VG+A+ FA A+ +A + +I+ +P+ID+ S+ +

Sbjct: 333 TVFFSVLIGAFSVGGASPSIEAFANARGAAYEIFKIIDNKPSIDSYKSGHK----- 384

Query: 11421 TTVDAK*QQAC*HTVLLCRLQEKYDGNVLFEDVKFNYPSPRPDVPVLQGLNLEVQKGETL 11600

+ GN+ F +V F+YPSR +V +L+GLNL+VQ G+T+

Sbjct: 385 -----PDNIKGNLEFRNVHFSYPSRKEVKILKGLNLKVQSGQTV 423

Query: 11601 ALVGSSGCGKSTTIQLLERFYDPREGRV 11684

ALVG+SGCGKSTT+QL++R YDP EG V

Sbjct: 424 ALVGNSSGCGKSTTVQLMORLYDPTEGMV 451

Score = 142 bits (354), Expect = 2e-32

Identities = 74/148 (50%), Positives = 99/148 (66%)

Query: 11244 VVMTMLYGAMAVGEANTYAPNFAKAKISASHLTMLINRQPAIDNLSEEEARLVNDTNPH 11423

V +++GAMAVG+ +++AP++AKAKISA+H+ M+I + P ID+ S E P+T

Sbjct: 977 VFSAVVFGAMAVGQVSSFAPDYAKAKISAHIIMIIEKTPLIDSYSTEGLM-----PNT 1030

Query: 11424 TVDAK*QQAC*HTVLLCRLQEKYDGNVLFEDVKFNYPSPRPDVPVLQGLNLEVQKGETLA 11603

+GNV F +V FNYP+RPD+PVLQGL+LEV+KG+TLA

Sbjct: 1031 L-----EGNVTFGVEVFNYPTRPDIPVLQGLSLEVKKGQTLA 1067

Query: 11604 LVGSSGCGKSTTIQLLERFYDPREGRVV 11687

LVGSSGCGKST +QLLERFYDP G+V+

Sbjct: 1068 LVGSSGCGKSTTVQLLERFYDPLAGKVL 1095

Score = 250 bits (632), Expect = 4e-65

Identities = 136/240 (56%), Positives = 159/240 (65%)

Query: 12136 LLDGVDVKQLNVHWLRSQIGIVSQEPVLFDCSLAENIAYGDNSRSVSMDEIVAAKAANI 12315
 LLDG ++K+LNV WLR+ +GIVSQEP+LFDCS+AENIAYGDNSR VS +EIV AAK ANI
 Sbjct: 1095 LLDGKEIKRLNVQWLRHLGIVSQEPILFDCSIAENIAYGDNSRVVSQEEIVRAAKEANI 1154
 Query: 12316 HSFIEGLPQVAAVNQKWLIPHLIDSHGAAHDHLHHIQTVSEQRVDTQAGDKGTQLSGGQ 12495
 H+FIE LP +Y T+ GDKGTQLSGGQ
 Sbjct: 1155 HAFIESLPN-----KYSTKVGDKGTQLSGGQ 1180
 Query: 12496 KQRVAIARAIIRNPKLLLLDEATSALDTESEKVRLYSH*CYVVHKAQAQHTSCVQVVQE 12675
 KQR+AIARA++R P +LLLDEATSALDTESEK VQE
 Sbjct: 1181 KQRIAIARALVRQPHILLDEATSALDTESEK-----VQE 1216
 Query: 12676 ALDQARKGRTCIVVAHRLSTIQNADCIAVFQGGVVVEKGTHTQQLIAKKGVYHMLVTKQMG 12855
 ALD+AR+GRTCIV+AHRLSTIQNAD I VFQ G V E GTHQQL+A+KG+Y +V+ Q G
 Sbjct: 1217 ALDKAREGRTCIVIAHRLSTIQNADLIVVFNQGRVKEHGTHQQLAQLKGIYFSMVSQAG 1276

Score = 196 bits (494), Expect = 6e-49

Identities = 114/237 (48%), Positives = 142/237 (59%)

Query: 12139 LDGVDVKQLNVHWLRSQIGIVSQEPVLFDCSLAENIAYGDNSRSVSMDEIVAAKAANI 12318
 +DG D++ +NV +LR IG+VSQEPVLF ++AENI YG +V+MDEI A K AN +
 Sbjct: 453 VDGQDIRTINVRFLREIIGVVSQEPVLFATTIAENIRYG---RENVTMDEIEKAVKEANAY 510
 Query: 12319 SFIEGLPQVAAVNQKWLIPHLIDSHGAAHDHLHHIQTVSEQRVDTQAGDKGTQLSGGQK 12498
 FI LP ++DT G++G QLSGGQK
 Sbjct: 511 DFIMKLPH-----KFDTLVGERGAQLSGGQK 536
 Query: 12499 QRVAIARAIIRNPKLLLLDEATSALDTESEKVRLYSH*CYVVHKAQAQHTSCVQVVQEA 12678
 QR+AIARA++RNP+LLLDEATSALDTESE V VQ A
 Sbjct: 537 QRIAIARALVRNPKILLDEATSALDTESEAV-----VQVA 572
 Query: 12679 LDQARKGRTCIVVAHRLSTIQNADCIAVFQGGVVVEKGTHTQQLIAKKGVYHMLVTKQ 12849
 LD+ARKGRT IV+AHRLST++NAD IA F GV+VEKG H +L+ +KG+Y LVT Q
 Sbjct: 573 LDKARKGRTTIVIAHRLSTVRNADVIAGFDDGVIVEKGNHDELMKEKGIYFKLVTMQ 629

12kb-26kb 的 blastx 结果

sp|P08183|MDR1_HUMAN MULTIDRUG RESISTANCE PROTEIN 1 (P-GLYCOPROTEIN 1)

Length = 1280

Score = 250 bits (632), Expect = 4e-65

Identities = 136/240 (56%), Positives = 159/240 (65%)

Query: 136 LLDGVDVKQLNVHWLRSQIGIVSQEPVLFDCSLAENIAYGDNSRSVSMDEIVAAKAANI 315
 LLDG ++K+LNV WLR+ +GIVSQEP+LFDCS+AENIAYGDNSR VS +EIV AAK ANI
 Sbjct: 1095 LLDGKEIKRLNVQWLRHLGIVSQEPILFDCSIAENIAYGDNSRVVSGEEIVRAAKEANI 1154
 Query: 316 HSFIEGLPQVAAVNQKWLIPHLIDSHGAAHDHLHHIQTVSEQRDYDTQAGDKGTQLSGGQ 495
 H+FI E LP +Y T+ GDKGTQLSGGQ
 Sbjct: 1155 HAFIESLPN-----KYSTKVGDKGTQLSGGQ 1180
 Query: 496 KQRVAIARAIIRNPKLLLLDEATSALDTESEKVRLYSH*CYVVHKAQAQHTSCVQVQVE 675
 KQR+AIARA++R P +LLLDEATSALDTESEK VQE
 Sbjct: 1181 KQRIATARALVRQPHILLDEATSALDTESEK-----VQE 1216
 Query: 676 ALDQARKGRTCIVVAHRLSTIQNADCIAVFQGGVVVEKGTHQQLIAKKGVYHMLVTKQMG 855
 ALD+AR+GRTCIV+AHRLSTIQNAD I VFQ G V E GTHQQL+A+KG+Y +V+ Q G
 Sbjct: 1217 ALDKAREGRTCIVIAHRLSTIQNADLIVVFNQGRVKEHGTHQQLLAQKGIYFSMVSQVQAG 1276

Score = 196 bits (494), Expect = 6e-49

Identities = 114/237 (48%), Positives = 142/237 (59%)

Query: 139 LDGVDVKQLNVHWLRSQIGIVSQEPVLFDCSLAENIAYGDNSRSVSMDEIVAAKAANI 318
 +DG D++ +NV +LR IG+VSQEPVLF ++AENI YG +V+MDEI A K AN +
 Sbjct: 453 VDGQDIRTINVRFLREIIGVVSQEPVLFATTIAENIRYG--RENVMTDEIEKAVKEANAY 510
 Query: 319 SFIEGLPQVAAVNQKWLIPHLIDSHGAAHDHLHHIQTVSEQRDYDTQAGDKGTQLSGGQK 498
 FI LP ++DT G++G QLSGGQK
 Sbjct: 511 DFIMKLP-----KFDLVGERGAQLSGGQK 536
 Query: 499 QRVAIARAIIRNPKLLLLDEATSALDTESEKVRLYSH*CYVVHKAQAQHTSCVQVQVEA 678
 QR+AIARA++RNPK+LLLDEATSALDTESE V VQ A
 Sbjct: 537 QRIATARALVRNPKIILLDEATSALDTESEAV-----VQVA 572
 Query: 679 LDQARKGRTCIVVAHRLSTIQNADCIAVFQGGVVVEKGTHQQLIAKKGVYHMLVTKQ 849
 LD+ARKGRT IV+AHRLST++NAD IA F GV+VEKG H +L+ +KG+Y LVT Q
 Sbjct: 573 LDKARKGRTTIVIAHRLSTVRNADVIAGFDDGVIVEKGNHDELMKEKGIYFKLVTMQ 629

Score = 96.7 bits (237), Expect(2) = 9e-31

Identities = 44/67 (65%), Positives = 54/67 (79%)

Query: 2937 RFAIYYSVLGFVLLAAYMQVSFWTITAGRQVKRIRSLFFHCIMQQEISWFDVNDTGELN 3116
 R+A YYS +G VL+AA+QVSFW + AGRQ+ +IR FFH IM+QEI WFDV+D GELN

Sbjct: 113 RYAYYYSGIGAGVLAAYIQVSWCLAAGRQIHKIRKQFFHAIMRQEIGWFDVHVDV GELN 172

Query: 3117 TRLTECV 3137

TRLT+ V

Sbjct: 173 TRLTDDV 179

Score = 60.9 bits (145), Expect(2) = $9e^{-31}$

Identities = 31/70 (44%), Positives = 40/70 (56%)

Query: 3233 DVYKIQEGIGDKVGLLIQAYXXXXXXXXXXXXXXXXWKLTLVILAVSPALISAFFSKVRG 3412

DV KI EGIGDK+G+ Q+ WKLTLVILA+SP L +SAA ++K+

Sbjct: 178 DVSKINEGIGDKIGMFFQSMATFFTFIVGFTRGWKLTIVILAI SPVLGLSAAVWAKILS 237

Query: 3413 SCLNPLLSLY 3442

S + L Y

Sbjct: 238 SFTDKELLAY 247

Score = 67.1 bits (161), Expect = $7e^{-10}$

Identities = 33/50 (66%), Positives = 40/50 (80%)

Query: 3459 LAVTSFGKVLASFTSKEQTAYAKAGAVAEVLSAIRTVFAFSGQTREIER 3608

L+ + K+L+SFT KE AYAKAGAVAEVL+AIRTV AF GQ +E+ER

Sbjct: 227 LSAAVWAKILSSFTDKELLAYAKAGAVAEVLA AIRTVIAFGGQKKELER 276

Score = 87.0 bits (212), Expect(2) = $6e^{-28}$

Identities = 38/59 (64%), Positives = 52/59 (87%)

Query: 3911 RYHKNLRDAKDVGVKKAISSNIAMGFTFLMIYLSYALAFWYGSTLILNFEYTIGNLLTV 4087

RY+KNL +AK +G+KKAI++NI++G FL+IY SYALAFWYG+TL+L+ EY+IG +LTV

Sbjct: 276 RYNKNLEEAKRIGIKKAITANISIGAAFLLIYASYALAFWYGTTLVLSGEYSIGQVLTV 334

Score = 61.3 bits (146), Expect(2) = $6e^{-28}$

Identities = 28/40 (70%), Positives = 35/40 (87%)

Query: 4165 IKVFFVVLIGAFSVGQTSFNQNFASARGAAYKVYSIIDN 4284

+ VFF VLIGAFSVGQ SP+I+ FA+ARGAAY+++ IIDN

Sbjct: 332 LTVFFSVLIGAFSVGQASPSIEAFANARGAAYEIFKIIDN 371

Score = 62.8 bits (150), Expect = $1e^{-08}$

Identities = 25/42 (59%), Positives = 35/42 (82%)

Query: 5284 VSQQPNIDSPSEDFKPDFIKGDIEFKNIHFNYPSRPEVKV 5409

+ KP+IDS+S+ G KPD IKG++EF+N+HF+YPSR EVK+
 Sbjct: 368 IIDNKPSIDSYSKSGHKPDNIKGNLEFRNVHFSYPSRKEVKI 409

Score = 44.5 bits (103), Expect(2) = 5e-17
 Identities = 16/42 (38%), Positives = 30/42 (71%)
 Query: 5284 VSQKPNIDSFSEDGFKPDFIKGDIEFKNIHFNYPSRPEVKV 5409
 + ++ P IDS+S +G P+ ++G++ F + FNYP+RP++ V
 Sbjct: 1011 IIEKTPLIDSYSTEGLMPNTLEGNVTFGEVVFNYPTRPDIPV 1052

Score = 66.7 bits (160), Expect(2) = 5e-17
 Identities = 30/45 (66%), Positives = 36/45 (79%)
 Query: 5469 LQILNMSLSVKSGQTIALVGSSGCGKSTTIQLLQRFYDPEEGAV 5603
 + +L +SL VK GQT+ALVGSSGCGKST +QLL+RFYDP G V
 Sbjct: 1050 IPVLQGLSLEVKKGQTLALVGSSGCGKSTVQLLERFYDPLAGKV 1094

Score = 188 bits (472), Expect = 2e-46
 Identities = 100/155 (64%), Positives = 114/155 (73%)
 Query: 5772 VFIDGHDIRSLNIRYLREMIGVVSQEPVLFATTITENIRYGRLDVTQEEIERATKESNAY 5951
 V +DG DIR++N+R+LRE+IGVVSQEPVLFATTI ENIRYGR +VT +EIE+A KE+NAY
 Sbjct: 451 VSDGQDIRTINVRFLREIIGVVSQEPVLFATTIAENIRYGRENTMDEIEKAVKEANAY 510
 Query: 5952 DFIMNLPDVRPHLWLPYLSLAPSRNANIYIMISEQKFETLVGDRGTQLSGGQKORIAIA 6131
 DFIM LP KF+TLVG+RG QLSGGQKORIAIA
 Sbjct: 511 DFIMKLP-----HKFDTLVGERGAQLSGGQKORIAIA 542
 Query: 6132 RALVRNPKILLLDEATSALDAESETIVQAALDKVQ 6236
 RALVRNPKILLLDEATSALD ESE +VQ ALDK +
 Sbjct: 543 RALVRNPKILLLDEATSALDTESEAVVQVALDKAR 577

Score = 146 bits (364), Expect = 1e-33
 Identities = 85/159 (53%), Positives = 105/159 (65%), Gaps = 3/159 (1%)
 Query: 5772 VFIDGHDIRSLNIRYLREMIGVVSQEPVLFATTITENIRYGRLD--VTQEEIERATKESN 5945
 V +DG +I+ LN+++LR +G+VSQEP+LF +I ENI YG V+QEEI RA KE+N
 Sbjct: 1094 VLLDGKEIKRLNVQWLAHLGIVSQEPILFDCSIAENIAYGDNRSRVVSQEEIVRAAKEAN 1153
 Query: 5946 AYDFIMNLPDVRPHLWLPYLSLAPSRNANIYIMISEQKFETLVGDRGTQLSGGQKORIA 6125
 + FI +LP+ K+ T VGD+GTQLSGGQKORIA
 Sbjct: 1154 IHAFIESLPN-----KYSTKVGDKGTQLSGGQKORIA 1185

Query: 6126 IARALVRNPKILLLDEATSALDAESETIVQAALDKV-QTQTC 6248

IARALVR P ILLLDEATSALD ESE +VQ ALDK + +TC

Sbjct: 1186 IARALVRQPHILLLDEATSALDTESEKVVQALDKAREGRTC 1227

Score = 80.8 bits (196), Expect = 5e-14

Identities = 37/55 (67%), Positives = 46/55 (83%)

Query: 6611 QVRLGRTTIVVAHRLSTIRNADIIAGFSNGKIVEQGTHSQLMEIKGVYHGLVTMQ 6775

+ R GRTTIV+AHRLST+RNAD+IAGF +G IVE+G H +LM+ KG+Y LVTMQ

Sbjct: 575 KARKGRTTIVIAHRLSTVRNADVIAGFDDGVIVEKGNHDELMKEKGIYFKLVTMQ 629

Score = 69.9 bits (168), Expect = 1e-10

Identities = 31/55 (56%), Positives = 42/55 (76%)

Query: 6611 QVRLGRTTIVVAHRLSTIRNADIIAGFSNGKIVEQGTHSQLMEIKGVYHGLVTMQ 6775

+ R GRT IV+AHRLSTI+NAD+I F NG++ E GTH QL+ KG+Y +V++Q

Sbjct: 1220 KAREGRTCIVIAHRLSTIQNADLIVVFQNGRVKEHGTHQQLLAQKGIYFSMVSQ 1274

Score = 73.4 bits (177), Expect(5) = 1e-62

Identities = 30/50 (60%), Positives = 40/50 (80%)

Query: 7419 DEDVDPVVSFFKVLHLNIPWPYILVGLICATINGAMQPVFAILFSKIITV 7568

DE +P VSF++++ LN+ EWPY +VG+ CA ING +QP FAI+FSKII V

Sbjct: 689 DESIPPVSFWRIMKLNLTWPYFVVGVFCAIINGGLQPAFAIIFSKIIGV 738

Score = 73.0 bits (176), Expect(5) = 1e-62

Identities = 52/169 (30%), Positives = 73/169 (42%), Gaps = 1/169 (0%)

Query: 7652 FAFLQVVFAD-PDRDSVRRKSEFISLMFVVIGCVSFTMFLQVSVFLTCICNAG*CLFES 7828

F+ ++ VF D ++ R+ S SL+F+ +G +SF+T FLQ

Sbjct: 732 FSKIIGVFTRIDDPETKRQNSNLFSLFLALGIISFITFFLQ----- 773

Query: 7829 NWPKNFQGGYCFGKSGEILTTLKRLRAFTAMMRQVSSDLNLNLTTRGFATATCPSGPAH 8008

G+ FGK+GEILT +LR F +M+RQ

Sbjct: 774 -----GFTFGKAGEILTKRLRYMVFRSMLRQ----- 799

Query: 8009 KYSFD*GFIKY*CKISILSVLKDLSWYDNPQNTVGALTTRLAADAAQVQG 8158

D+SW+D+P+NT GALTTRLA DAAQV+G

Sbjct: 800 -----DVSWFDDPKNTTGALTTRLANDAAQVKG 827

Score = 62.5 bits (149), Expect(5) = 1e-62

Identities = 32/75 (42%), Positives = 39/75 (51%)
 Query: 8232 AAGVRLATIMQNFANLGTSTIIIAFVYGWXXXXXXXXXXXXXXXXXXXXXXXXXGHAAKDK 8411
 A G RLA I QN ANLGT III+F+YGW+ G A KDK
 Sbjct: 828 AIGSRLAVITQNIANLGTGIIISFIYGWQLTLLLAIVPIIATAGVEMKMLSGQALKDK 887
 Query: 8412 KELEKAGKVELKRVK 8456
 KELE AGK+ + ++
 Sbjct: 888 KELEGAGKIAIEAIE 902

Score = 50.4 bits (118), Expect(5) = 1e-62
 Identities = 23/35 (65%), Positives = 31/35 (87%)
 Query: 8519 QIATEAIENVRTVVSLSREPKFECLYEENLRVPYK 8623
 +IATEAIEN RTVVSL++E KFE +Y ++L+VPY+
 Sbjct: 895 KIATEAIENFRVVSLSLQEQKFEHMYAQLQVPYR 929

Score = 67.5 bits (162), Expect(5) = 1e-62
 Identities = 30/54 (55%), Positives = 40/54 (73%)
 Query: 8698 SLDFADRNSQKKAHVYGLTYSFSQAMIYFAYAACFRFGAWLIEAGRMDVEGVFL 8859
 SL RNS +KAH++G+T+SF+QAM+YF+YA CFRFGA+L+ M E V L
 Sbjct: 923 SLQVPYRNSLRKAHIFGITFSFTQAMMYFSYAGCFRFGAYLVAHKLMSFEDVLL 976

Score = 41.0 bits (94), Expect(2) = 1e-23
 Identities = 21/54 (38%), Positives = 33/54 (60%)
 Query: 9567 VVSAVLYGAMAVGEANTFAPNYAKAKMAASYLMLINKKPAIDNLSEEGTSPVN 9728
 V +VL GA +VG+A+ +A A+ AA + +I+ KP+ID+ S+ G P N
 Sbjct: 334 VFPSVLIGAFPSVQASPSIEAFANARGAAYEIFKIIDNKPSIDSYKSGHKPDN 387

Score = 92.4 bits (226), Expect(2) = 1e-23
 Identities = 41/62 (66%), Positives = 52/62 (83%)
 Query: 9818 GNVHFEGVKFNYPSPDVTILQGLNLKVKKGETLALVSGCGKSTTIQLLERFYDPREG 9997
 GN+ F V F+YPSR +V IL+GLNLKV+ G+T+ALVG+SGCGKSTT+QL++R YDP EG
 Sbjct: 390 GNLEFRNVHFSYPSRKEVKILKGLNLKVGSGQTVALVNSGCGKSTTVQLMORLYDPTEG 449
 Query: 9998 RV 10003
 V
 Sbjct: 450 MV 451

Score = 65.6 bits (157), Expect(2) = 1e-33

Identities = 29/52 (55%), Positives = 43/52 (81%)

Query: 9567 VVSAVLYGAMAVGEANTFAPNYAKAKMAASYLMLINKKPAIDLSEEGTSP 9722

V SAV++GAMAVG+ ++FAP+YAKAK++A+++M+I K P ID+ S EG P

Sbjct: 977 VFSAVVFGAMAVGQVSSFAPDYAKAKISAAHIIMIIEKTPILDSYSTEGLMP 1028

Score = 101 bits (249), Expect(2) = 1e-33

Identities = 46/64 (71%), Positives = 56/64 (86%)

Query: 9815 DGNVHFEGVKFNYPSPDVTILQGLNLKVKKGETLALVGSSGCGKSTTIQLLERFYDPRE 9994

+GNV F V FNYP+RPD+ +LQGL+L+VKKG+TLALVGSSGCGKST +QLLERFYDP

Sbjct: 1032 EGNVTFGEVFNYPTRPDIPVLQGLSLEVKKGQTLALVGSSGCGKSTVVQLLERFYDPLA 1091

Query: 9995 GRVV 10006

G+V+

Sbjct: 1092 GKVL 1095

Score = 54.7 bits (129), Expect = 4e-06

Identities = 26/50 (52%), Positives = 38/50 (76%)

Query: 10825 LDGVNVKQLNIHWLRSQIGIVSQEPVLFDCSLAENIAYGDNSRSVSMDEI 10974

+DG +++ +N+ +LR IG+VSQEPVLF ++AENI YG +V+MDEI

Sbjct: 453 VDGQDIRTINVRFLREIIGVVSQEPVLFATTIAENIRYG--RENVTMDEI 500

Score = 84.7 bits (206), Expect(2) = 4e-59

Identities = 37/51 (72%), Positives = 45/51 (87%)

Query: 10825 LDGVNVKQLNIHWLRSQIGIVSQEPVLFDCSLAENIAYGDNSRSVSMDEIV 10977

LDG +K+LN+ WLR+ +GIVSQEP+LFDCS+AENIAYGDNSR VS +EIV

Sbjct: 1096 LDGKEIKRLNVQWLRHLGIVSQEPILFDCSIAENIAYGDNSRVVSQEEIV 1146

Score = 168 bits (420), Expect(2) = 4e-59

Identities = 97/187 (51%), Positives = 112/187 (59%)

Query: 10982 AAKAANIHSFIEGLPQVAAVNQKWLIPHLIDSXXXXXXXXXXXXTVSEQRDYDTQAGDKG 11161

AAK ANIH+FIE LP +Y T+ GDKG

Sbjct: 1148 AAKEANIHFIESLPN-----KYSTKVGDKG 1173

Query: 11162 TQLSGGQKQRAIARAIRNPKLLLLDEATSALDTESEKVRLYSH+CYVVHKAQAQHCCTS 11341

TQLSGGQKQRAIARA++R P +LLLDEATSALDTESEKV

Sbjct: 1174 TQLSGGQKQRIAIARALVRQPHILLLDEATSALDTESEKV----- 1213

Query: 11342 CFQVVQEALDQARKGRTICIVVAHRLSTIQNADCIAVFQGGVVVEKGTHQQLIAKKGVYHM 11521
 VQEALD+AR+GRTCIV+AHRLSTIQNAD I VFQ G V E GTHQQL+A+KG+Y
 Sbjct: 1214 ———VQEALDKAREGRTICIVIAHRLSTIQNADLIVVFQNGRVKEHGTHQQLLAQKGIYFS 1269
 Query: 11522 LVTKQMG 11542
 +V+ Q G
 Sbjct: 1270 MVSQAG 1276

Score = 148 bits (371), Expect = 2e-34

Identities = 78/135 (57%), Positives = 95/135 (69%)

Query: 11132 RYDTQAGDKGTQLSGGQKQRIARAIIRNPKLLLLDEATSALDTESEKVRLYSH*CYVV 11311
 ++DT G++G QLSGGQKQR+AIARA++RNPK+LLLDEATSALDTESE V
 Sbjct: 519 KFDTLVGERGAQLSGGQKQRIARALVRNPKLLLLDEATSALDTESEAV————— 568
 Query: 11312 HKAAAQHCTSCFQVVQEALDQARKGRTICIVVAHRLSTIQNADCIAVFQGGVVVEKGTHQQ 11491
 VQ ALD+ARKGRT IV+AHRLST++NAD IA F GV+VEKG H +
 Sbjct: 569 —————VQVALDKARKGRTTIVIAHRLSTVRNADVIAGFDDGVIVEKGNHDE 614
 Query: 11492 LIAKKGVYHMLVTKQ 11536
 L+ +KG+Y LVT Q
 Sbjct: 615 LMKEKGIYFKLVTMQ 629

三、使用剪接位点预测软件 (splice site prediction) (网址: <http://www-hgc.lbl.gov/projects/splic.html>)来预测 1bp-14kb 和 12kb-26kb 中可能存在的剪接位点, 结果如下:

Splice site predictions for 1 sequence with donor score cutoff 0.00, acceptor score cutoff 0.00 (exon/intron boundary shown in larger font):

Donor site predictions for 1bp-14kb :

Start	End	Score	Exon	Intron
78	92	0.08	caggctggtgataag	
97	111	0.00	cacagctgacagggg	
178	192	0.00	aggccacgtctgctgc	
198	212	0.13	actgagagtgagaag	
206	220	0.92	tgagaaggtgagact	
278	292	0.08	tgtccaggtggtgca	
292	306	0.00	aggagcgttgacc	
319	333	0.01	gcaggacgtgcatcg	
395	409	0.00	agtgggtggtgaaaa	

439	453	0.01	agggagtgtaccaca
452	466	0.00	catgctggtcaccaa
468	482	0.00	cagatgggtcaacac
491	505	0.12	aaagacggtgacaac
546	560	0.24	gtgttgtgtaagact
689	703	0.02	agctcaggtcccgtt
796	810	0.83	gataaaggtgtggat
824	838	0.00	gctttgggtgacaga
883	897	0.02	ttggcaggtcatttg
915	929	0.00	attacaagtcacacc
933	947	0.01	gcgctacgtgatttc
1008	1022	0.01	tcaccacgtgtcttt
1042	1056	0.01	agctgaggtccataa
1074	1088	0.00	cagacatgtagtgca
1132	1146	0.99	ctcatgggtgagtg
1175	1189	0.00	catgaccgttggctc
1212	1226	0.01	gcccagtgaacaca
1221	1235	0.13	aacacatgtaacaac
1276	1290	0.01	gtgtttcgtgaggtc
1293	1307	0.00	catcatggttctggg
1316	1330	0.01	cagtcacgtctgagg
1344	1358	0.70	catcaaagtaaatgc
1366	1380	0.04	atcagcgggtgcctct
1417	1431	0.08	ccccaatgtcagagg
1486	1500	0.03	aaataaggtcgttga
1514	1528	0.01	ggtttgggtgatact
1588	1602	0.09	atagaacgttagagt
1610	1624	0.00	aatctgggtccagca
1620	1634	0.00	cagcagggttttag
1645	1659	0.00	tgcaacagtcagatc
1662	1676	0.00	gctgcatgtgcacgg
1689	1703	0.00	taacagggtccccta
1700	1714	0.05	cctaaaggttgaggg
1718	1732	0.00	aagagatgtttgaat
1729	1743	0.01	gaatgaagtagaatc

1744	1758	0.75	agactaagtaaataa
1785	1799	0.90	agagcaggtactcac
1824	1838	0.15	caaataagtgagcgt
1965	1979	0.02	tgtcagggtcattaa
2146	2160	0.99	atttgaggtaggtct
2228	2242	0.00	aagaaatgttatatt
2273	2287	0.02	tttatatgtatatat
2281	2295	0.01	tatatatgtacacag
2300	2314	0.00	ggctcctgtagtgac
2320	2334	0.01	aaataatgtagagga
2374	2388	0.97	cacattggtaagatc
2557	2571	0.01	ccgggagtagcatt
2597	2611	0.01	tgcacacgtgcacgt
2605	2619	0.00	tgcacgtgtagttcc
2614	2628	0.01	agttccggttgcag
2625	2639	0.01	gcagatggtgggac
2681	2695	0.16	cgggctggtgaatcc
2709	2723	0.23	gtgtttggtgagatg
2775	2789	0.97	aaccccagtgagtgt
2804	2818	0.42	gcatcaagtgagaca
2817	2831	0.28	cacacacgtaaaaaa
2885	2899	0.99	tgcagaggtacgttc
3017	3031	0.68	ctcacaggtatttga
3047	3061	0.01	tgtcatggttcattt
3094	3108	0.01	catagaggtctcgct
3110	3124	0.00	aggaaacgttacatt
3137	3151	0.00	gcaacatgttccaat
3162	3176	0.06	aaacctggtgtctgg
3184	3198	0.00	catagtggttctaga
3215	3229	0.00	tgttgtggttcttca
3301	3315	0.02	aggtcatgtaccata
3349	3363	0.00	tcagcaggtctccaa
3371	3385	0.02	ggctcctgtgcgttc
3436	3450	0.00	gagtttggttactct
3494	3508	0.00	tcttggggtttgctg

3507	3521	0.04	tgtgctggtagtggc
3515	3529	0.01	tagtggcgtagctgc
3614	3628	0.75	tcagctggtagacg
3663	3677	0.01	agagtgggtgacgca
3726	3740	0.00	atgtgggttcgatt
3739	3753	0.84	ttccctggtagcgac
3775	3789	0.02	ggcatcggtagacaag
3787	3801	0.00	aaggcgggtctgctg
3864	3878	0.01	caccctggtagatcct
3918	3932	1.00	cagtaaggtaaataat
3941	3955	0.00	accagtggtagccata
3999	4013	0.06	ctgcctggtagggaca
4145	4159	0.02	ttcaccagtaaagag
4161	4175	0.15	agacagcgtagacca
4192	4206	0.03	agcggaggtgctatc
4215	4229	0.00	ggactgtgtttgcct
4251	4265	1.00	tcaaaaggtgagtgt
4305	4319	0.01	ggactgagtcagagt
4321	4335	0.60	ggagcaggttaatat
4487	4501	0.77	ccttctggtagcgga
4517	4531	0.01	acaaaggtacacca
4542	4556	0.84	actgactgtgagtgg
4611	4625	0.82	tgaccacgtgtgtcc
4627	4641	0.02	aatgaaggtgttctt
4741	4755	0.99	cgaccacgtgagtcc
4753	4767	0.06	tccagtggtagtgaat
4888	4902	0.01	ccagctggtagcag
4906	4920	0.06	cattttggtagttta
4943	4957	0.28	agttactgttagtac
4957	4971	0.04	cctgttggtagtttc
5043	5057	0.04	tttgcacgtgaacac
5052	5066	0.24	gaacacagtagcac
5103	5117	0.02	ccaagggttctttg
5112	5126	0.00	tctttgggttcaaca
5130	5144	0.02	tctaaacgtagagat

5147	5161	0.98	tagagcggtaactcc
5176	5190	0.00	tcattctgtaaaatt
5351	5365	0.00	attggttgtagttc
5379	5393	0.00	ccgtcgggtccaaca
5497	5511	0.01	agtactagtactgct
5656	5670	0.29	atcaaaggtgacatt
5712	5726	0.96	aattaaagtatgttc
5822	5836	0.30	gtttcatgtgaggaa
5849	5863	0.02	tgctttggtggggag
5868	5882	0.11	ggctgtggtaaaagt
5899	5913	0.08	tgcagaggttctacg
5924	5938	0.97	aggatccgtaagtaa
5948	5962	0.64	tttgaagtatgttt
5984	5998	0.01	tggcagggtattgaa
6002	6016	0.84	attccaggtgtgaga
6039	6053	0.40	attgacagtaggtgc
6051	6065	0.00	tgacatgtgcacta
6070	6084	0.04	gagagaggtgaagct
6081	6095	0.00	agcttgagtgatgct
6146	6160	0.01	tcaaaatgtgctttt
6206	6220	0.01	ctgcctagtgaatgc
6228	6242	0.03	ttaacatgtgaaaac
6240	6254	0.38	aacacacgtgtgtag
6314	6328	0.05	cggagtggtcagcca
6380	6394	0.01	actggacgtgacgca
6452	6466	0.76	tccagacgtgaggcc
6538	6552	0.01	agcagaagtttgaga
6551	6565	0.05	gacgctggtgggaga
6617	6631	0.00	agctctggtccgcaa
6655	6669	0.00	aggccacgtctgccc
6686	6700	0.00	gaccatcgtacaggc
6707	6721	0.81	ggacaaggtacagac
6723	6737	0.01	cagacacgtgcagag
6791	6805	0.03	tttcatggtgtctgg
6936	6950	0.60	ttctcccgtgagggc

6956	6970	0.02	ttttaaggttgtttt
6983	6997	0.00	ttttcaggtctccca
7008	7022	0.67	catccaggtctgctt
7084	7098	0.81	tctacaggtccgact
7169	7183	0.02	agtaatggtgaaatc
7234	7248	0.01	tggcctggtgacat
7246	7260	0.83	catgcaggtatccaa
7271	7285	0.64	agtcgacgtgcgttt
7298	7312	0.00	gagtgtagtcagaac
7444	7458	0.01	ggaggaggtccacta
7507	7521	0.01	aatttgagtgcgatc
7532	7546	0.87	agaggaggtgagctg
7542	7556	0.00	agctgtggtccagag
7592	7606	0.03	cttaacagtgagata
7625	7639	0.02	tttataggtctggtt
7699	7713	0.21	gttactggtactctt
7712	7726	0.09	tttcccagtggttaa
7755	7769	0.00	tgagaacgttccctcc
7779	7793	0.01	ctttaaagtgatgcg
7808	7822	0.00	agtgccgatalattt
7818	7832	0.61	tattttggttaggaac
7846	7860	0.68	atcaacggtgcgatg
7893	7907	0.05	cattatggtgcatta
7936	7950	0.00	tttcagcgtttgtca
7972	7986	0.00	atrtcaggtcttcag
8047	8061	0.00	gatgggggtggtgac
8077	8091	0.86	tctgcaggtttgtat
8112	8126	0.00	ctttgatgtgtgctg
8161	8175	0.00	gccccgggttgcca
8181	8195	0.00	agcagcggtttagtc
8200	8214	0.78	tgcaaaggtcagatg
8216	8230	0.54	agttttggtaaacac
8241	8255	0.01	tgctattgtcagtcc
8250	8264	0.01	cagtcctgtcagcac
8337	8351	0.00	tggttgtgtagcact

8351	8365	0.01	ttactaggtgttggga
8425	8439	0.98	ccatgatgtgagtct
8455	8469	0.61	cccaaaggtgctatc
8476	8490	0.12	agatctgggtgactgt
8491	8505	0.02	ggaggccgtagagt
8502	8516	0.01	gagtgcagtcagttt
8535	8549	0.01	atatgaggtctttga
8564	8578	0.02	gctggacgtagccat
8579	8593	0.00	cgctgtggtcataaa
8590	8604	0.03	taaaggggtgaaatg
8616	8630	0.55	cactcaagtaggctg
8648	8662	0.02	tcagttggtactaaa
8766	8780	0.02	ctggaaagtagtttt
8776	8790	0.14	gttttatgtaaatga
8857	8871	0.67	agtgagggtaaacag
8868	8882	0.04	acagtaggtctggag
8914	8928	0.30	aattgaggttaaaag
8980	8994	0.04	gcagcatgtaaaaat
9120	9134	0.01	ataagcagtacttac
9212	9226	0.01	ttccagggtttctgt
9224	9238	0.06	tgtttcagtaaatct
9289	9303	0.99	gagacaggtgagtga
9358	9372	0.07	tatat ttgtaaatat
9366	9380	0.01	taaatatgtatctcg
9423	9437	0.34	taataaggtgccgac
9533	9547	0.07	tattgatgtaaaatc
9573	9587	0.74	tcagctggtacgaca
9637	9651	0.40	cgcccacgtacaagg
9646	9660	0.99	acaaggagtaagtct
9664	9678	0.00	caggacgtgacatt
9728	9742	0.13	tcaggggtgagcct
9740	9754	0.00	cctggctgtaatgac
9793	9807	0.07	gcttcgtgtacggct
9830	9844	0.00	ggccgtggtgcccat
9863	9877	0.07	cgctgaggtcaagct

9923	9937	0.14	cggaaaggttttct
9967	9981	0.01	gggggatgtgagatg
10010	10024	0.00	ctgcgtggtttctcc
10048	10062	0.14	cgagaatgtgagaac
10060	10074	0.00	aactgtggtgtccct
10125	10139	0.61	catacaagtaggacc
10363	10377	0.94	tgttctctgtgagtac
10392	10406	0.01	ctatactgtaaagac
10443	10457	0.00	atcccaagttacaac
10458	10472	0.06	accaacagtaaagac
10505	10519	0.14	tctaaagttatggt
10555	10569	0.00	tgagactgtgggggc
10603	10617	0.01	tcttaatgtagcacc
10622	10636	0.00	acccatggttggcat
10750	10764	0.15	gtataatgtacatag
10762	10776	0.11	tagtaatgtacataa
10823	10837	0.08	gtaacacgtgaattt
10848	10862	0.00	gaaatcagtaaagtt
10894	10908	0.01	gaagcgtgttagcgc
10906	10920	0.36	cgcgtcggtagatg
10954	10968	0.01	tctaaaagttaaact
11009	11023	0.93	gcacaccgtaagtga
11033	11047	0.04	ttgtctggttaatga
11063	11077	0.02	aagtcacgtttgagt
11138	11152	0.08	ggattgtgtaaataa
11170	11184	0.03	aattcaggttgagaa
11185	11199	0.04	aacgcaagtcagatc
11240	11254	0.00	cagtgtggttatgac
11254	11268	0.02	caatgctgtacggcg
11393	11407	0.84	gagactggtaaataa
11459	11473	0.00	acacacggttcttct
11498	11512	0.00	cggcaacggttctttt
11513	11527	0.00	tgaggacgtcaagtt
11549	11563	0.00	tgtgcctgtactaca
11573	11587	0.09	tctggaggtgcaaaa

11771	11785	0.02	atccaaagtagctta
11805	11819	0.00	attaacagtcagcgt
11850	11864	0.05	ttaagaggttcaaag
11915	11929	0.00	agcatctgtgggaca
12000	12014	0.00	ccccacagtatatga
12048	12062	0.18	ccacaaggtccatgt
12072	12086	0.00	gggcaaagtgagcgc
12087	12101	0.00	tttactcgtaccatt
12139	12153	0.00	ctggacggtgtcgat
12147	12161	0.00	tgtcgatgtgaaaca
12176	12190	0.00	ggctgaggtctcaga
12207	12221	0.02	ggagccggtgctgtt
12336	12350	0.56	gcctcaggtagcggc
12362	12376	0.00	ggaaatggttgattc
12457	12471	0.08	caggctggtgataag
12476	12490	0.00	cacagctgtcagggg
12557	12571	0.00	aggccacgtctgcgc
12577	12591	0.13	actgagagtgagaag
12585	12599	0.92	tgagaaggtgagact
12657	12671	0.08	tgtccaggtggtgca
12671	12685	0.00	aggagcgttgacc
12698	12712	0.01	gcaggacgtgcatcg
12774	12788	0.00	agtgggtggtgaaaa
12818	12832	0.01	agggagtgtaccaca
12831	12845	0.00	catgctggtcaccaa
12900	12914	0.00	cagacctgtatcaaa
12976	12990	0.02	tcggcatgtacagtt
13028	13042	0.95	aattcagtaactta
13138	13152	0.01	tacgcatgtactcat
13149	13163	0.01	tcatcatgtactgat

13315	13329	0.93	aaataaagtaaact
13327	13341	0.06	tctatatgtgagga
13382	13396	0.29	cgttagcgtgagaat
13396	13410	0.03	tctagcggttgatc
13424	13438	0.00	agattccgtgcagct
13445	13459	0.94	gatccacgtgaggag
13480	13494	0.01	cagacgcgtgtgata
13551	13565	1.00	actaacggtacgttc
13594	13608	0.11	ctaaccggtgtgtgt
13616	13630	0.00	tactcgggtgtcatc
13646	13660	0.15	tgatgacgtaacact
13662	13676	0.00	aatgagcgttcgcgg
13768	13782	0.09	acaaacggtgatctg
13778	13792	0.95	atctgagglaaaaat
13803	13817	0.01	aataaaagtagtaat
13830	13844	0.00	ttaatacgttaaatt
13848	13862	0.42	taaaactgtgcgtaa
13912	13926	0.03	ttatittgtaaaat
13944	13958	0.00	gacgatggtgccaag

Acceptor site predictions for 1bp-14kb :

Start	End	Score	Intron	Exon
45	85	0.10	atacatactgtctctgagcagagatacgcactcaggctgg	
115	155	0.00	agaagcagcgtgtcgccatagcccagccatcatccgcaac	
213	253	0.30	gtgagactttattcacattagtgtctacgtagtcataaggc	
223	263	0.04	attcacattagtgtctacgtagtcataaggccgctgcacaa	
231	271	0.00	tagtgtctacgtagtcataaggccgctgcacaacactgcac	
264	304	0.85	cactgcacctcttgtgtccagggtggtgcaggaggcgttgga	
273	313	0.01	tcttgtgtccagggtggtgcaggaggcgttgaccaggccag	
319	359	0.02	gcaggacgtgcatcgtcgtagcccaccgtctgtccaccatc	
342	382	0.79	caccgtctgtccaccatccagaacgccgactgcatcgtgt	
369	409	0.49	gactgcatcgtgtgttccaggaggagtggtggtggaaaa	
450	490	0.02	cacatgctggtcaccaaacagatgggtcaacacagtggata	
517	557	0.04	catttttacatcgcaatccagtttaatgggtgttgtgtaag	

619	659	0.62	tatttaaccattgcttcatagttcattatcgtgatcaatct
649	689	0.97	gtgatcaatcttctcttttagctcagacgactcctcagtca
666	706	0.00	ttagctcagacgactcctcagtcagctcaggtcccgttcat
675	715	0.18	acgactcctcagtcagctcaggtcccgttcatcaccattt
765	805	0.49	tctctgtcttcttttctgaggagctctgatgataaagggtg
802	842	0.01	gggtgtggataaacttgctcaagggtttgggtgacagatggt
817	857	0.02	ctcaagggtttgggtgacagatggttcttctcttttctca
838	878	0.92	atgttgcctcgttttctcagtatgtttaaagaacaat
851	891	0.12	tttctcagtatgtttaaagaacaatgaacttggcaggt
869	909	0.01	aagaacaatgaacttggcaggtcatttgttaaatttgaag
934	974	0.84	cgctacgtgatttcttggcagatcatttttgggaggcaagt
985	1025	0.00	aatttgcctcattatgcaaagattcaccacgtgtctttta
1056	1096	0.61	aaccttcatatgccattcagacatgtagtgcagagaaggc
1069	1109	0.01	ccattcagacatgtagtgcagagaaggccttagaacctca
1109	1149	0.10	atlttacatctaactactagaactcatgggtgagtgtccc
1176	1216	0.03	atgaccgttggctccctaagctcacatactaagatgcca
1219	1259	0.00	gtaacacatgtaacaacacagatctttgataaaaataaaga
1295	1335	0.35	tcatggttctgggttcttccagcagtcacgtctgaggaggga
1408	1448	0.00	tcccataatccccaatgtcagaggcagctgtcaagcttcat
1432	1472	0.00	cagctgtcaagcttcatgcagccaaggcagcagaacaacct
1486	1526	0.01	aaataaggctgtgacgatagaacgagcggtttgggtgata
1517	1557	0.11	ttgggtgatactctaataccagactggagccgttcaggaatg
1571	1611	0.38	aaacttaacctgatgtatagaacgttagagtgaacaggaa
1579	1619	0.00	cctgatgtatagaacgttagagtgaacaggaatctgggtc
1588	1628	0.00	atagaacgttagagtgaacaggaatctgggtccagcagggt
1605	1645	0.00	acaggaatctgggtccagcagggttttaggaacaaccttt
1614	1654	0.21	tgggtccagcagggttttaggaacaacctttgcaacagtc
1636	1676	0.00	aacaacctttgcaacagtcagatctggctgcatgtgcacgg
1674	1714	0.39	cggttgtgcttcagataacagggtcccctaaaggttgaggg
1686	1726	0.04	agataacagggtcccctaaaggttgagggccgaagagatgt
1766	1806	0.41	aactactttatagattacagagcaggtactcaccacacag
1786	1826	0.48	gagcaggtactcaccacacaggactgacaagtcaaagacaa
1831	1871	0.67	gtgagcgtttcccttcttaagattatagcttcttgggact
1839	1879	0.01	ttcccttcttaagattatagcttcttgggactcatcccta
1874	1914	0.05	tcctacaaccactactgcaggggatgacgccctcttgaat

1917	1957	0.24	tctcattttctatctacgcagccttagtgtttctctgttt
1950	1990	0.53	cctgttttataatttgtcagggcattaaaaatgatccac
2001	2041	0.18	ttgaagcgttctgtcttaagatccctgggatttttacttg
2059	2099	0.08	aaaataaatctaaactcttagatgccagcaggcactcctga
2069	2109	0.02	taaactcttagatgccagcaggcactcctgaaaggacaaa
2177	2217	0.02	atattgatgtgttgacttagaaaaataagaggaattaac
2275	2315	0.01	tatatgtatatgtacacagttatggctcctgtagtgaca
2290	2330	0.03	acacagttatggctcctgtagtgacaacaaaaataatgtag
2365	2405	0.00	tggccccatcacattggaagatcagccccctgctctgtgat
2406	2446	0.00	aattgataaataagtttttagcttgacttattgtagttgtg
2421	2461	0.45	tttagcttgacttattgtagttgtgatgccgttcttctgc
2535	2575	0.34	aaaattattactccacatagccccgggagtagcatttaac
2595	2635	0.00	tgtgcacacgtgcacgtgtagttccggtttgcatggttg
2608	2648	0.43	acgtgtagttccggtttgcatggttgggacatcttgatg
2721	2761	0.01	atgactgacagcttcatccaggaagccaaactgtcccaaaa
2776	2816	0.02	accccagtgagtgctcctcagctgaggagcatcaagtgaga
2837	2877	0.92	taaacttggttttactctaggagcaaacagcaccttagaa
2915	2955	0.00	aacaatgtctaattgtttaagagtgctgcatcattcgcc
2943	2983	0.81	tgcatcattcgccatctgtagataatgacaggggagcaatt
2953	2993	0.00	gccatctgtagataatgacaggggagcaattgttaaatcag
3003	3043	0.47	ttttcaataaaacctcacaggtatttgacgcctgaggagc
3078	3118	0.01	ccttttgactggaaaccatagaggctcgcctcaggaaacgt
3168	3208	0.25	ggtgtctggccccatcatagtggttctagatatattcagc
3177	3217	0.12	ccccatcatagtggttctagatatattcagcttggcctgt
3187	3227	0.00	agtggttctagatatattcagcttggcctgttgggttctt
3210	3250	0.79	tggcctgttgggttcttctcagctcggactgatgtccagga
3228	3268	0.00	cagctcggactgatgtccaggaaagaaactatactataaa
3260	3300	0.00	tactataaaactatactatagtcacctatgaaccatattt
3282	3322	0.07	ccacctatgaaccatatttaggtcatgtaccataatcctca
3324	3364	0.01	cctctctcacacatcaccagcctctcagcaggctccaac
3335	3375	0.50	acatcaccagcctctcagcaggctccaaccatcggctc
3383	3423	0.08	ttccgacatcttgattgcttagtcgataaaggctcgtcgct
3392	3432	0.00	cttgattgcttagtcgataaaggctcgtcgctccttaccta
3413	3453	0.07	gctcgtcgctccttacctagccgagtttggttactcta
3453	3493	0.88	tctcaacctctattttttagattctccatctactactcca

3497	3537	0.06	tggggtttgctgtgctggtagtggcgtacctgcagatgtct
3511	3551	0.03	ctggtagtggcgtacctgcagatgtctctgtggaccctaac
3544	3584	0.00	accctaacggccgcgcggcaggccaaacgaattcgcgagtt
3589	3629	0.07	ttccacggcatcatgcagcaggacatcagctggtatgacgt
3644	3684	0.04	agctcaacacgcgtctcacagagtgggtgacgcacatcata
3674	3714	0.05	cgcacatcatacacactccagttcctgtcacagctggcgtg
3686	3726	0.59	aacctccagttcctgtcacagctggcgtggtcttatcata
3751	3791	0.00	cacgatgtctacaagatccaggagggcatcggtgacaaggc
3787	3827	0.02	aaggcgggtctgctgatccaggcggcctccacctttatcac
3867	3907	0.04	cctggtcacctggccatcagccctgtgttgggtctctcag
3887	3927	0.38	gccctgtgttgggtctctcagctgccctttacagtaaggta
3900	3940	0.87	tctctcagctgccctttacagtaaggtaaataatctccatac
3941	3981	0.01	accagtggtgccataatccagtcctcgagggccacagccca
3957	3997	0.01	tccagtcctcgagggccacagcccagccaaaggaagcgttt
3994	4034	0.01	gttttctgcctgggtgggacagaaaaccccggtggggaccac
4045	4085	0.03	tttgggcacccccatccatagacacgattactcccctctta
4112	4152	0.54	cctcaacctgtggccccatagttgctgacaagcttcaccag
4195	4235	0.50	ggaggtgctatcctccatcaggactgtgtttgccttcagtg
4213	4253	0.83	caggactgtgtttgccttcagtgccaaagaaaagccatca
4265	4305	0.33	ttctaattgattattaccttagattatcttacatttatggg
4320	4360	0.01	gggagcaggttaataattccagacatcgctaactttagtttt
4336	4376	0.02	tccagacatcgctaactttagttttcctttgtgttcctcag
4356	4396	0.96	gttttcctttgtgttcctcagatacataagaacctggagg
4427	4467	0.04	ttgctgctaacacggccacaggcttctcctttctgatgatc
4502	4542	0.01	ctactctggtcctcaacaaagagtacaccattggaaattha
4565	4605	0.01	atcatcacttccaccattaagaataagagcgttgctgcaga
4584	4624	0.00	agaataagagcgttgctgcagaaacagtgaccacgtgtgtc
4655	4695	0.04	ggggcatacatatttggacaggcctctcccaacgtccagtc
4673	4713	0.02	caggcctctccaacgtccagtcctttgccagtgccagagg
4690	4730	0.23	ccagtcctttgccagtgccagaggagcggcgtataaaagtct
4807	4847	0.03	aaacacgcacgatgtcatagggacagctagctgtgctcat
4907	4947	0.87	atthtggatthtaatttcaggctgggctgtttacagtha
4924	4964	0.17	tcaggctgggctgtttacagttactgttagtacctgttgg
4934	4974	0.00	cctgtttacagttactgttagtacctgttggatthtcaca
4979	5019	0.01	aacatathaaatgactttagcgaagcgaatgctatcactt

5039	5079	0.00	tctgtttgcacgtgaacacagtatgcacacagcttagagcc
5055	5095	0.00	cacagtatgcacacagcttagagccaactggaaaaaggcat
5088	5128	0.00	aaaggcattaaatctccaaagggttctttgggttcaacatc
5120	5160	0.00	ttcaacatcgtctaaacgtagagattttagagcggtaactc
5129	5169	0.00	gtctaaacgtagagattttagagcggtaactcctctgcttt
5181	5221	0.38	ctgtaaaattttgtgtttaaggcttataaaatgaaatcgct
5239	5279	0.92	gttccactcttgccttcatagtttcttgagctctctttttt
5281	5321	0.99	tttgggtgtttctctttccaggaattgctttgtagttcctg
5295	5335	0.17	tttccaggaattgctttgtagttcctgtttctgtctctgt
5342	5382	0.12	cttgcctgcgattgggtgttagttcaagagcttgccagccgt
5439	5479	0.01	agctggctgacacacattcagatgccagtgtagcctcatgc
5451	5491	0.00	cacattcagatgccagtgtagcctcatgctagcctcactgc
5462	5502	0.00	gccagtgtagcctcatgctagcctcactgctggacagtact
5478	5518	0.00	gctagcctcactgctggacagtactagtactgctggatagc
5565	5605	0.00	gaagccttcgacttctaaagcatttctgccatatctcaac
5587	5627	0.10	atttctgccatatctcaacagaaaccgaatattgacagctt
5612	5652	0.08	cgaatattgacagcttttcagaggacggatacaagcctgaa
5659	5699	0.03	aaaggtgacattgtattccagaacatccacttcagctacc
5673	5713	0.05	attccagaacatccacttcagctacccttcgaggccagaaa
5685	5725	0.00	ccacttcagctacccttcgaggccagaaattaaagtatggt
5720	5760	0.71	tatgttcaccatccatcacagttttctatatggaaacaatg
5784	5824	0.93	actcactctctcaccatccagatcttaaacgacatgctggt
5883	5923	0.01	accaccatccagctgctgcagaggttctacgacccccagaa
5901	5941	0.01	cagaggttctacgacccccagaaaggatccgtaagtaattt
5935	5975	0.02	gtaatttgggtctattttgaagtatgtttgaagattttcc
5947	5987	0.01	attttgaagtatgttttgaagatttttctgagcatttggc
5969	6009	0.21	tttttctgagcatttggcagggtattgaatttattccagg
5988	6028	0.56	agggtattgaatttattccaggttgagaaaggacatgatg
6029	6069	0.02	tcctcaaccattgacagtaggtgcacatgtgcactatata
6050	6090	0.20	gtgcacatgtgcactatatagagagaggtgaagcttgagtg
6111	6151	0.21	agtcacccaaacttcattcagatttttataaaacatcaaaa
6234	6274	0.00	tgtgaaaacacacgtgtgtagatattatcgacggtcacga
6436	6476	0.11	acttcatcatgaaccttcagacgtgaggcctcatttatgg
6522	6562	0.01	tacattatgatctctgagcagaagtttgagacgctgggtggg
6678	6718	0.01	gagagcgagaccatcgtacaggtgctctggacaaggtaca

6715	6755	0.00	tacagaccagacacgtgcagagatgatgggaaacaaagaa
6805	6845	0.06	gaaaagactcccagcccacagaagccctcccagagcccca
6817	6857	0.02	agcccacagaagccctcccagagcccccagagctccaaatt
6826	6866	0.00	aagccctcccagagcccccagagctccaaattgcccctact
6850	6890	0.01	tccaaattgcccctactaaagcatgcagcaggatcaaatag
6942	6982	0.14	cgtgagggcatttgttttaaggttgttttttagtcagctta
6953	6993	0.20	ttgttttaaggttgttttttagtcagcttattttcaggtct
6969	7009	0.89	tttagtcagcttattttcaggtctcccagcttctaattca
6978	7018	0.19	gcttattttcaggtctcccagcttctaattcatccaggtct
6994	7034	0.56	cccagcttctaattcatccaggtctgcttgagttctgtact
7070	7110	0.99	taaaatgggttttctctacaggtccgactgggtcgcaccac
7123	7163	0.04	tcaccgactctcgaccatcagaaacgccgacatcattgctg
7150	7190	0.05	cgacatcattgctggattcagtaatggtgaaatcgttgagc
7232	7272	0.23	catggcctggtgaccatgcaggtatccaaaactcagaggag
7285	7325	0.02	tctgtcgtactctgagtgtagtcagaacgtccatcctggga
7323	7363	0.95	ggacatcttttctgtctacagagctttcagaagctggagga
7332	7372	0.01	ttcctgtctacagagctttcagaagctggaggatctggaaga
7410	7450	0.89	ctgatcgaatccttctcccagctctccctgcagaggaggag
7422	7462	0.86	ttctcccagctctccctgcagaggaggaggtccactagagg
7462	7502	0.60	gctccttgcttgctgtctcagaaggaacaaaaggagagaaa
7570	7610	0.11	ggcggctcgccctcagattcagacttaacagtgagataagcc
7611	7651	0.79	aggatcagctgacctttataggtctggttatagttgctgggt
7623	7663	0.16	cctttataggtctggttatagttgctggctctaaacttct
7658	7698	0.00	actttccatacatttaagcagcgtttttagtgcatgtgtt
7668	7708	0.19	catttaagcagcgtttttagtgcatgtgttgttactggta
7699	7739	0.42	gttactgggtactcttcccagtggtaacacagcttccaat
7711	7751	0.01	cttcccagtggtaacacagcttccaataatctttctgca
7732	7772	0.86	cttccaataatcttctgcaggtgagaacgttccctcccg
7766	7806	0.35	ctcccgtgtcgttctttaaagtgatgcgttacaacgttct
7808	7848	0.78	agtggccgtatattttggttaggaacctctgcgcatgatc

8063	8103	0.96	tttctccatgtttctgcaggttgtatccactgaatccc
8122	8162	0.02	tgctgttttacattttgaaaggaaaaagcatagaaaaagc
8186	8226	0.03	cggttagctcaatgcaaaggtcagatgtagttttggtaa
8232	8272	0.03	tggattctatgctattgtcagtcctgtcagcacatttctag
8241	8281	0.04	tgctattgtcagtcctgtcagcacatttctagaggcttcac
8252	8292	0.03	gtcctgtcagcacatttctagaggcttcactcacctagtc
8270	8310	0.04	tagaggcttcactcaccttagctttcaacaccacctaag
8327	8367	0.04	cacaaattgttggttgtgtagcacttactaggtgttgaaa
8337	8377	0.19	tggttgtgtagcacttactaggtgttgaaatgttctctgg
8384	8424	0.01	tggccattttgacatgatagcatcagcatcagctgcacat
8432	8472	0.77	gtgagtctccctttccagcagatcccaaaggtgctatcgga
8441	8481	0.02	cctttccagcagatcccaaaggtgctatcggattaagatct
8511	8551	0.55	cagtttatgtcatgtccagaagatatgaggtctttgatg
8554	8594	0.00	gcattattctgctggacgtagccatcgcctgtggtcataaag
8574	8614	0.01	gccatcgcctgtggtcataaaggggtaaatgatcattaacg
8613	8653	0.00	cgacactcaagtagctgcagcattcaaatgatgctcagtt
8643	8683	0.00	gatgctcagttggtactaaagggtctttctttttcaactg
8691	8731	0.00	gtgtgctgcgtttgggtccagctccatttgtatctttcac
8790	8830	0.02	acatttgatcctggccggcagcaaccgtttacaagcctct
8830	8870	0.35	taaacttgtcttcccactaaggcttgaagtgagggtaaaca
8900	8940	0.00	tgctgaactctttaattgaggttaaaagtacctccagcc
8927	8967	0.02	agttacctccagcctccaagggtttacctgctacccc
8956	8996	0.89	acctgctaccctctaccagatggcagcatgtaaaaaataa
9044	9084	0.66	tttaaccactccttcatcagagagaataagatcgcagcat
9120	9160	0.03	ataagcagtacttaccttagactactataaatgatcattt
9182	9222	0.89	catcacatctctttaaacagttttggggtttccagggttt
9197	9237	0.46	aaacagttttggggtttccagggtttctgtttcagtaaatc
9211	9251	0.31	tttccagggtttctgtttcagtaaatcggagaaattctga
9244	9284	0.00	aattctgacctgaacctcagactcaaagccttcatatcta
9275	9315	0.01	ttcatatctatgatgagacaggtgagtgagataaaccttaa
9326	9366	0.12	tttgggattcttttgaacaggagcatttgatatatttgt
9380	9420	0.01	ggggcactggactttaaatagataatgcaaatcttggatg
9449	9489	0.21	cttttatcccacacatgcagatcatgcccgatcagcaac
9548	9588	0.42	tcaatcctctctgtattgaaggacctcagctggtagcacia
9601	9641	0.38	cgttggcgtctcaccactaggctggctgccgacgccccc

9667	9707	0.01	ggacgtgacatttgacctcagcgaggcctaattaacggagg
9705	9745	0.92	aggatgccccctctctgctcaggctgcaggggtgcgcctggc
9738	9778	0.01	cgcctggctgtaatgacgcagaacttcgccaacctgggcac
9773	9813	0.00	gggcaccagcatcatcatcagcttcgtgtacggctgggagc
9865	9905	0.02	ctgaggtcaagctgctgacaggacacgccgccgaagacaag
10010	10050	0.99	ctgcgtggtttctccctgcagatgccacagaggccatcga
10020	10060	0.07	tctccctgcagatgccacagaggccatcgagaatgtgaga
10060	10100	0.25	aactgtgggtgcccctcaccagagaaccgacatttgtggctt
10115	10155	0.01	ctaactgttccatacaagtaggaccaatgacaccattcctc
10147	10187	0.01	ccattcctcatatgtcgacagtatttcatgcattttatgt
10173	10213	0.02	tcatgcattttatgtatccagtgtttctcatcacttttt
10208	10248	0.90	ttttttctcactctgcgaggaaactcccagaaaaaggcca
10260	10300	0.89	ttaacctactccttctcacaggccatgatctctttgttta
10303	10343	0.03	ctgcctgtttccgctttggagcctggctgatcgaagcagga
10407	10447	0.55	tcaatgcttttatttccaaaggacagcagagcagacatccc
10481	10521	0.01	gaggagcttcatcttctaaagtcttctaaaggttatgttag
10491	10531	0.28	atcttctaaagtcttctaaaggttatgttagccaagctgac
10593	10633	0.00	catccacaactcttaatgtagccaataaccatggttgg
10685	10725	0.52	acattaattttcttcatatagtctaatattcccttctgtat
10778	10818	0.95	aaaatagttttttcttttagtgtgtatattactggctac
10856	10896	0.69	taaagttttatcttcttagtctcttgagtcacgattgaa
10865	10905	0.00	atcttctttagtctcttgagtcacgattgaagcgtttag
10910	10950	0.00	tcggtgagatgacttcccagcatgccttggcacagcttgt
10925	10965	0.02	tcccagcatgccttggcacagcttgttttctaaaagttaa
10941	10981	0.02	cacagcttgttttctaaaagttaaacttaattctattaaa
11042	11082	0.01	taatgagtggcttgcacccaagaagtcacgtttgagtttaatt
11074	11114	0.02	gagttaattaatgctgtatagtgtttgcagtgcagagttac
11083	11123	0.03	aatgctgtatagtgtttgcagtgcagagttaccctgtcatt
11115	11155	0.90	cctgtcattgttctatttttagcgggattgtgtaataattt
11156	11196	0.00	ggagcctaagattaaattcaggttgagaaaacgcaagtcag
11222	11262	0.45	gtcattaaccttatttcccagtggttatgacaatgctgt
11296	11336	0.00	atgctcccaacttcgcaaagccaaaatctcagcctccac
11350	11390	0.04	taaacagacagccggccatagataatctgtcagaggaggaa
11362	11402	0.00	cggccatagataatctgtcagaggaggaaagcagactggta
11466	11506	0.95	gttcttctctgctctccaggagaaatcagcggcaacgt

11496	11536	0.00	gacggcaacgttctttttgaggacgtcaagtttaactacc
11545	11585	0.00	ccgatgtgcctgtactacaagggctgaatctggaggtgcaa
11597	11637	0.00	tctggccttggtgggcagcagcggttgaggaaagacacca
11651	11691	0.04	ggagaggttttatgaccccagagaggggagagtggtagaa
11694	11734	0.50	caacagtcattatcttcttagtgtctctatgtaaaccagtc
11740	11780	0.01	aacaccttttgctcatgaagcattctagaaatccaaagta
11748	11788	0.01	ttgcctcatgaagcattctagaaatccaaagtagcttatct
11777	11817	0.16	agtagcttatctaactgttagctaatgaattaacagtcagc
11896	11936	0.01	agaggagcctggctaccagcatctgtgggacatccagat
11914	11954	0.02	cagcatctgtgggacatccagataatgtcccaggctgagct
11926	11966	0.01	gacatccagataatgtcccaggctgagcttgagacctcag
11955	11995	0.02	tggagacctcagccagctcaggagctctcatgctttctcaa
11987	12027	0.79	ctttctcaacctccccacagtatatgagagaagggacatg
12026	12066	0.02	tgatcatcggccttctccaagtcacaaggtccatggtgat
12034	12074	0.01	ggccttctccaagtcacaaggtccatggtgatatttaggg
12052	12092	0.17	aaggtccatggtgatatttagggcaaagtgagcggttact
12115	12155	0.86	atcacatttgtgctctctcagttgctggacggtgtcgatgt
12169	12209	0.09	gttactggctgaggtctcagatcggcatcgtctcccagga
12187	12227	0.43	cagatcggcatcgtctcccaggagccggtgctgttcgactg
12311	12351	0.04	acatccacagcttcatcgaagggctgcctcaggtagcggct
12322	12362	0.36	ttcatcgaagggctgcctcaggtagcggctgtgaatcaggg
12424	12464	0.04	atacaactgtctctgagcagagatacgcactcaggctgg
12494	12534	0.00	agaagcagcgtgtcgccatagcccagccatcatccgcaac
12592	12632	0.21	gtgagactttattcacactagtgtctacgtagtcataaggc
12602	12642	0.03	attcacactagtgtctacgtagtcataaggccgctgcacaa
12610	12650	0.00	tagtgtctacgtagtcataaggccgctgcacaacactgcac
12643	12683	0.85	cactgcacctcttgtgtccaggtggtgcaggaggcgttggaa
12652	12692	0.01	tcttgtgtccaggtggtgcaggaggcgttggaccaggccag
12698	12738	0.02	gcaggacgtgcacgtctgtagccaccgctgtccaccatc
12721	12761	0.79	caccgtctgtccaccatccagaacgccgactgcatcgtgt
12748	12788	0.49	gactgcatcgtgtgttccagggaggagtggtggtggaaaa
12829	12869	0.02	cacatgctggtcacciaacagatgggctatcacagtggatg
12895	12935	0.00	ggcagcagacctgtatcaagacttctcaacagttggagc
12908	12948	0.12	tatcaagacttctcaacagttggagcctattcaggactt
12923	12963	0.02	caacagttggagcctattcaggactttatttattgttact

12950	12990	0.05	atztatgttgactgctgaagactgtcggcatgtacagtt
12968	13008	0.01	aagacttgctggcatgtacagttcccatacgtcacccctcag
12988	13028	0.01	gttcccatacgtcacccctcagcatccctgagtgtttgttaa
13014	13054	0.29	ctgagtgtttgttaaattcagtaacttattttatgccatg
13055	13095	0.04	ctggagtgtttgttaacagcaggaataggaatgaaaaataa
13160	13200	0.88	tgatggctgtggtttcttagatcagttttgattttacata
13257	13297	0.91	agcaataaacctcctttcagatagatagatagatatggc
13265	13305	0.02	accctcctttcagatagatagatagatatggccacgcccc
13405	13445	0.01	ttgatcgggatcgactcagattccgtgcagctgttaatg
13416	13456	0.01	tcgcactcagattccgtgcagctgttaatgatccacgtgag
13495	13535	0.01	accgccctctgtccaaaatagaccagctcgcagggtctctg
13508	13548	0.05	caaaatagaccagctcgcagggtctcgggacaccggcagc
13592	13632	0.10	ttctaaccctgtgtgttcagttgtactcgggtgtcatctg
13709	13749	0.87	taacgcgttttctgttgacaggactgtcaatcatggcctta
13732	13772	0.07	ctgtcaatcatggccttaaagatcgatacggccgaaacaaa
13880	13920	0.00	tttggttacaacacttaatagtcgccatttcgttatttgt
13910	13950	0.01	cgttatttgtaaatatgtagccatgattccaaggacgatg
13923	13963	0.00	atatgtagccatgattccaaggacgatggtgccaagaatga

Splice site predictions for 1 sequence with donor score cutoff 0.00, acceptor score cutoff 0.00 (exon/intron boundary shown in larger font):

Donor site predictions for 12kb-26kb :

Start	End	Score	Exon	Intron
48	62	0.18	ccacaaggtccatgt	
72	86	0.00	gggcaaagtggagcg	
87	101	0.00	tttactcgtaccatt	
139	153	0.00	ctggacgggtgcgat	
147	161	0.00	tgtcgatgtgaaaca	
176	190	0.00	ggctgaggtctcaga	
207	221	0.02	ggagccggtgctgtt	
336	350	0.56	gcctcaggtagcggc	
362	376	0.00	ggaaatggttgattc	
457	471	0.08	caggctggtgataag	

476	490	0.00	cacagctgtcagggg
557	571	0.00	aggccacgtctgcgc
577	591	0.13	actgagagtgagaag
585	599	0.92	tgagaaggtgagact
657	671	0.08	tgtccaggtggtgca
671	685	0.00	aggaggcgttgacc
698	712	0.01	gcaggacgtgcatcg
774	788	0.00	agtgggtggtgaaaa
818	832	0.01	agggagtgtaccaca
831	845	0.00	catgctggtcaccaa
900	914	0.00	cagacctgtatcaaa
976	990	0.02	tcggcatgtacagtt
1028	1042	0.95	aattcagtaactta
1138	1152	0.01	tacgcatgtactcat
1149	1163	0.01	tcatcatgtactgat
1315	1329	0.93	aaataaagtaaact
1327	1341	0.06	tctatatgtgagggg
1382	1396	0.29	cgtagcgtgagaat
1396	1410	0.03	tctagcggtttgatc
1424	1438	0.00	agattccgtgcagct
1445	1459	0.94	gatccacgtgaggag
1480	1494	0.01	cagacgcgtgtgata
1551	1565	1.00	actaacggtacgttc
1594	1608	0.11	ctaaccggtgtgtgt
1616	1630	0.00	tactcgggtgtcatc
1646	1660	0.15	tgatgacgtaacact
1662	1676	0.00	aatgagcgttcgcgg
1768	1782	0.09	acaaacggtgatctg
1778	1792	0.95	atctgaggtaaaaat
1803	1817	0.01	aataaaagtagtaat
1830	1844	0.00	ttaatacgttaaatt
1848	1862	0.42	taaaactgtgcgtaa
1912	1926	0.03	ttattingtaaatat
1944	1958	0.00	gacgatggtgccaag
2003	2017	0.03	gcccattggtgggccc

2021	2035	0.67	tactctggtaataac
2098	2112	0.87	tctttaggtatatct
2106	2120	0.10	tatatctgtaaataa
2196	2210	0.03	agaaactgtgtgtgt
2206	2220	0.00	tgtgtgtgtgtgtgt
2210	2224	0.00	tgtgtgtgtgtgtgt
2218	2232	0.00	tgtgtgtgtgtgtgt
2220	2234	0.00	tgtgtgtgtgtgtgt
2222	2236	0.00	tgtgtgtgtgtgtgt
2230	2244	0.00	tgtgtgtgtgtgtgt
2240	2254	0.00	tgtgtgtgtgtgtgt
2248	2262	0.00	tgtgtgtgtgtgtac
2257	2271	0.14	gtgtacagtaaataa
2268	2282	0.30	atgaggggtaaagga
2484	2498	0.00	ggccatgggtcaacgg
2496	2510	0.00	cggcaccgtgatgcc
2510	2524	0.02	ccctcatgtgcatgt
2575	2589	0.01	aacgcaagtggctgg
2584	2598	0.20	ggctggagtgagttt
2682	2696	0.00	acagcacgttacagg
2703	2717	0.89	tgcaaaggtgtgttt
2729	2743	0.06	tcttactgtaacttt
2811	2825	0.01	cactcctgtgatgtt
2895	2909	0.01	ttaacaggtctttgg
2990	3004	0.05	catgcaggtgtcctt
3023	3037	0.60	gcgccaggtgaaacg
3079	3093	0.01	tcagctggtttgacg
3124	3138	0.12	tgacggagtgcgtat
3164	3178	0.03	acgctctgtacggct
3182	3196	0.01	ctcggaggtgtagat
3202	3216	0.01	gatggacgtgcttct
3254	3268	0.02	ggcatcgggtgacaag
3262	3276	0.61	tgacaaggtggggct
3282	3296	0.01	tccaggcgtacacca
3343	3357	0.00	gacgctgggtcatcct

3355	3369	0.39	cctggccgtgagccc
3397	3411	0.92	cagtaaagtaagagg
3476	3490	0.30	tgggaaggtgcttgc
3495	3509	0.86	ttcaccagtaaggag
3511	3525	0.08	agacggcgtaccca
3542	3556	0.00	ggaggaagtgctgtc
3565	3579	0.01	ggaccgtgttcgcct
3579	3593	0.23	ttcagtggtcagacc
3601	3615	0.21	ttgagaggtgcttac
3625	3639	0.04	tttgcacgtgaaagc
3703	3717	0.00	ttcattagtaatgtg
3711	3725	0.04	taatgtggtttgact
3727	3741	0.06	aaagacggtattgca
3739	3753	0.05	gcagggggtgaatat
3784	3798	0.07	aaaagctgtaactgg
3892	3906	0.01	gttctgcgtgtgtcg
3937	3951	0.03	aaaggacgtgggagt
4023	4037	0.77	ccttctggtacggga
4031	4045	0.06	tacgggagtacgctc
4053	4067	0.00	atthttgagtacacca
4078	4092	0.87	actgactgtgagtac
4118	4132	0.14	gagggtggtgagggt
4148	4162	0.01	tgcacttgtgtgtct
4164	4178	0.17	cattaaggtgttttt
4259	4273	0.00	ataaagtgtacagca
4278	4292	0.98	cgataacgtaaggag
4358	4372	0.02	gttctgtgttagtga
4374	4388	0.01	tgcgcgcgtcagcac
4397	4411	0.00	tctgggagtccgctt
4417	4431	0.61	cagagaggtgggggg
4455	4469	0.00	tgtttgggttcattt
4468	4482	0.02	ttcacaggtcacttg
4567	4581	0.92	tcaacaggtgaatag
4611	4625	0.02	cgagaaagtgtgacg
4656	4670	0.44	agcgcacgtgagctc

4720	4734	0.00	cctcactgtcagccg
4737	4751	0.01	aggaagtgttaggtc
4789	4803	0.18	ttactttgtaaggaa
4859	4873	0.01	aattcacgtggctta
4900	4914	0.01	gcagagggtctgggc
4922	4936	0.01	tacgctgtgtgtaa
5000	5014	0.02	tttgatggtttgcct
5097	5111	0.00	acactttgtgtgttg
5154	5168	0.00	agccccgtgaaacc
5249	5263	0.01	ttctcgggttaaaaa
5344	5358	0.24	atcaaaggtgacatc
5400	5414	0.58	agtcaaagtggtgtc
5440	5454	0.00	agactgggtctgatg
5519	5533	0.02	tgctttggtggggag
5544	5558	0.04	ggcaaaagtaccacc
5569	5583	0.08	tgcagaggttctacg
5594	5608	0.69	aggagctgtgagtca
5710	5724	0.97	cattctggtaaatat
5721	5735	0.00	atattcagtaacta
5737	5751	0.05	tagtgaggtgatgct
5765	5779	0.44	tgtgtaggtatttat
5831	5845	0.05	cggagtggtcagcca
5897	5911	0.01	actggacgtgacgca
5969	5983	0.76	tccagacgtgaggcc
6055	6069	0.01	agcagaagttgaga
6068	6082	0.05	gacgctggtgggaga
6134	6148	0.00	agctctggtccgcaa
6172	6186	0.00	aggccacgltgccc
6203	6217	0.00	gaccatcgtacaggc
6224	6238	0.81	ggacaaggtacagac
6255	6269	0.45	tctccatgtaggacc
6292	6306	0.00	tttctcagtatgggc
6301	6315	0.06	atgggcagtatgtgt
6317	6331	0.00	tagtggggtgttaag
6356	6370	0.00	tcagaatgtgacgtc

6388	6402	0.03	acaggacgtatttat
6396	6410	0.14	tatttatgtaactgt
6442	6456	0.00	gtcaaaggctctacac
6528	6542	0.17	atcaaaggctcaacat
6548	6562	0.28	cgctagcgtgaggaa
6596	6610	0.00	ggtttttgtaaactc
6607	6621	0.68	actccaggctccgact
6686	6700	0.00	ggattcagtaatggc
6757	6771	0.01	tggcctggtgaccat
6769	6783	0.32	catgcaggtgacccc
6807	6821	0.93	agttaaagtatgttt
6895	6909	0.00	ccgcatgtcggagt
6979	6993	0.00	ccagagggtcctcgt
6994	7008	0.00	ttgccgcgtcagaag
7028	7042	0.11	agaagagggtccgtg
7084	7098	0.11	agttaaagtaatatt
7220	7234	0.55	cagacagggtgctaca
7233	7247	0.63	cactcagggtgataac
7248	7262	0.02	atc gatggttattac
7298	7312	0.02	cctgagggtgcatcg
7350	7364	0.11	aatagaagtgaatca
7395	7409	0.00	aggctccgtgcgctg
7403	7417	0.00	tgcgctggttccgca
7484	7498	0.01	catccttgtggggct
7559	7573	0.78	catcactgtaggtcc
7579	7593	0.00	gctgaccgtgggacg
7600	7614	0.01	tcaactgggtcaaatg
7663	7677	0.04	attgcagggtgttgc
7682	7696	0.01	ccagaccgtgattct
7690	7704	0.00	tgattctgtcaggag
7722	7736	0.01	ctctgatglttgtcg
7739	7753	0.00	attggctgtgtgtca
7768	7782	1.00	tttacaggttaagtgt
7793	7807	0.00	tgtatttghtaatgct
7847	7861	0.14	caacagggttactgt

7859	7873	0.98	tgtttcggtaaacc
7924	7938	0.97	gagacaggtgagctc
8021	8035	0.00	gattgaggtttcatc
8039	8053	0.07	tattgatgtaaaac
8079	8093	0.74	tcagctggtagaca
8143	8157	0.33	cgccaagtacaagg
8152	8166	0.99	acaaggagtaagtct
8170	8184	0.00	cagggacgtgacatt
8234	8248	0.13	tgcaggggtgagcct
8336	8350	0.00	ggccgtggtgcccct
8385	8399	0.00	ctcgcgggtcacgcc
8429	8443	0.03	cggaaaggtggagtt
8496	8510	0.02	aataaaagtgtggtt
8539	8553	0.03	cgagaacgtcagaac
8616	8630	0.45	cgtacaagtatgaca
8735	8749	0.05	cccacgtgtacggct
8852	8866	0.94	tgttctgtgagtac
8889	8903	0.01	ggcttctgtatgtg
8900	8914	0.08	gttgaaggtgacatt
8928	8942	0.00	aggagacgtgcaaag
8950	8964	0.00	cagggcagtaaagtg
8977	8991	0.00	ctgaacagtttgtgg
9004	9018	0.45	gaccaggtcaggac
9053	9067	0.01	cacacatgtgttct
9105	9119	0.08	atataaagtaacata
9128	9142	0.84	aaccgaggtattacc
9241	9255	0.06	ggagttgtaataat
9276	9290	0.03	aatctaggtggaaga
9303	9317	0.16	ataat tggtaacaaat
9368	9382	0.00	aagcactgtacaac
9507	9521	0.03	caatagggtttgata
9560	9574	0.02	tcccaggtggtttc
9577	9591	0.01	cggtgctgtacggcg
9593	9607	0.00	catggccgtggggga
9716	9730	0.98	gtctccgtaaacac

9742	9756	0.56	cacaaaggtgacaaa
9812	9826	0.66	tacgacggtaatgtg
9862	9876	0.00	gcccgatgtgaccat
9892	9906	0.57	cctgaaggtgaaaa
9919	9933	0.01	ggccttggtgggcag
9969	9983	0.00	tggagaggttttatg
9997	10011	0.82	gagagtggtgagaag
10024	10038	0.00	tgctgtggtttttt
10059	10073	0.00	aacccccgtgcctct
10076	10090	0.12	acctactgttagttc
10099	10113	0.25	ttagctcgtatgtgt
10164	10178	0.02	ctgtgcagtacggat
10172	10186	0.00	tacggatgttgttc
10303	10317	0.01	cattcttgaacgtc
10343	10357	0.01	actggtggctctgtt
10381	10395	0.05	gacaaatgtagctgc
10409	10423	0.00	cgcaactgtagaagc
10469	10483	0.05	catttatgtaaacctc
10510	10524	0.00	ttggcaggttgagaa
10575	10589	0.00	ttgcagcgtgggcgc
10712	10726	0.04	tgatgcagtatgact
10795	10809	0.00	agctaaagttctcct
10825	10839	0.00	ctggacgggtcaac
10833	10847	0.02	tgtcaacgtgaaaca
10862	10876	0.00	ggctgaggctcaga
10893	10907	0.02	ggagccggtgctgtt
11023	11037	0.56	gcctcaggtagcggc
11049	11063	0.00	ggaaatggttgattc
11144	11158	0.08	caggctggtgataag
11163	11177	0.00	cacagctgtcagggg
11244	11258	0.00	aggccacgtctgcgc
11264	11278	0.13	actgagagtgagaag
11272	11286	0.92	tgagaaggtgagact
11344	11358	0.08	ttccaggtggtgca
11358	11372	0.00	aggaggcgttgacc

11385	11399	0.01	gcaggacgtgcatcg
11395	11409	0.02	catcgtggtagccca
11461	11475	0.00	agtgggtggtgaaaa
11505	11519	0.01	agggagtgtaccaca
11518	11532	0.00	catgctggtcaccaa
11602	11616	0.03	ctgaaatgtaattta
11752	11766	0.03	aacacacgtgcagtg
11766	11780	0.00	gactgaagttggcag
11799	11813	0.01	cgctgcggtgcctgt
11839	11853	0.02	tctttagttagtgt
11886	11900	0.00	ggttgaagtctgaac
11910	11924	0.00	gaaaaatgtatcctg
11954	11968	0.48	attcagagtgaggac
12025	12039	0.02	agagaatgtgggggt
12086	12100	0.00	acgttctcgtgaatca
12116	12130	0.02	actaacgtgaccaa
12150	12164	0.68	aggaccagtaagatg
12166	12180	0.01	agcgaacgtgccacc
12283	12297	0.56	atgtacagtaaatac
12490	12504	0.17	agggaaaggtttgcat
12507	12521	0.00	atcagaggttttgtt
12525	12539	0.03	tgtggaggtgctcag
12553	12567	0.27	ttgccaggtgcagca
12608	12622	0.01	ggcgtcggctctggcg
12656	12670	0.01	cctgcgggtcaaagt
12688	12702	0.08	gaatgtggtgggtca
12721	12735	0.02	gtgatccgtcagtag
12752	12766	0.05	ggatacggtagaaga
12770	12784	0.00	cgtagccgtggggaa
12807	12821	0.00	aactgtggtgtagcc
12904	12918	1.00	ataaaaggtgagtga
12957	12971	0.00	ccacgcagtacctct
13041	13055	0.66	cgagaaggtgcctgt
13077	13091	0.47	cgacgacgtaacagg
13125	13139	0.02	agacaaagtgaaggg

13134	13148	0.00	gaagggcgtgactca
13171	13185	0.10	ccatcaggttgttgt
13301	13315	0.02	acgacgggtccgtca
13338	13352	0.02	ctccacgggtgcaggg
13370	13384	0.06	tgccgtggtagaact
13442	13456	0.68	attccgggtcagccc
13461	13475	0.35	ccccgggtgcgttt
13492	13506	0.00	acctctggtgcagtt
13504	13518	0.44	gtttaagtagcca
13543	13557	0.03	tgtccgggtgcagct
13583	13597	0.00	gaaagccgtgaaggc
13591	13605	0.02	tgaaggcgtaacaga
13609	13623	0.03	cctgcaggtcctgcg
13699	13713	0.00	actcaaagtattgtt
13748	13762	0.02	gtccacagtaaacac
13815	13829	0.00	tcacaatgtaatatag
13862	13876	0.29	gggaatggtgactac
13886	13900	0.02	gcctccggtgctttg
13909	13923	0.20	ggtccacgtaccagc
13942	13956	0.01	tggcgtcgtacggcg

Acceptor site predictions for 12kb-26kb :

Start	End	Score	Intron	Exon
26	66	0.02	tgatcatcggccttctccaagtcacaaaggtccatggtgat	
34	74	0.01	ggccttctccaagtcacaaaggtccatggtgatatttaggg	
52	92	0.17	aaggtccatggtgatatttagggcaaagtgagcgtttact	
115	155	0.86	atcacatttgtgctctctcagttgctggacggtgctgatgt	
169	209	0.09	gttactggctgaggtctcagatcggcatcgtctcccagga	
187	227	0.43	cagatcggcatcgtctcccaggagccggtgctgttcgactg	
311	351	0.04	acatccacagcttcatcgaagggtgcctcaggtagcggct	
322	362	0.36	ttcatcgaagggtgcctcaggtagcggctgtgaatcaggg	
424	464	0.04	atacaaactgtctctgagcagagatacgacactcaggtgg	
494	534	0.00	agaagcagcgtgtcgccatagcccagccatcatccgcaac	
592	632	0.21	gtgagactttattcacactagtgtacgtagtcataaaggc	

602	642	0.03	attcacactagtgtacgtagtcataaggccgctgcacaa
610	650	0.00	tagtgctacgtagtcataaggccgctgcacaactgcac
643	683	0.85	cactgcacctcttgtgtccaggtggtgcaggaggcgttga
652	692	0.01	tcttgtgtccaggtggtgcaggaggcgttggaccaggccag
698	738	0.02	gcaggacgtgcatcgctcgtagcccaccgtctgtccaccatc
721	761	0.79	caccgtctgtccaccatccagaacgccgactgcatcgctgt
748	788	0.49	gactgcatcgctgtgtccagggaggagtggagggtgaaaa
829	869	0.02	cacatgctggtcaccaaacagatgggctatcacagtggatg
895	935	0.00	ggcagcagacctgtatcaaagacttcctcaacagttggagc
908	948	0.12	tatcaaagacttcctcaacagttggagcctattcaggactt
923	963	0.02	caacagttggagcctattcaggactttatttattgttgact
950	990	0.05	atttattgttgactgctgaagacttgctggcatgtacagtt
968	1008	0.01	aagacttgtcggcatgtacagttcccatacgtcacctcag
988	1028	0.01	gtcccatacgtcacctcagcatccctgagtgtttgttaa
1014	1054	0.29	ctgagtgtttgttaaattcaggtaacttatttattgccatg
1055	1095	0.04	ctggagtgttgtaacagcaggaataggaaatgaaaaataa
1160	1200	0.88	tgatggctgtggttttcttagatcagttttgatttacata
1257	1297	0.91	agcaataaacctcctttcagatagatagatagatattggc
1265	1305	0.02	accctcctttcagatagatagatagatattggccacgccca
1405	1445	0.01	ttgatcggggatcgactcagattccgtgcagctgttaatg
1416	1456	0.01	tcgactcagattccgtgcagctgttaatgatccacgtgag
1495	1535	0.01	accgccctctgtccaaaatagaccagctcgcagggtctg
1508	1548	0.05	caaaaatagaccagctcgcagggtctgggacaccggcagc
1592	1632	0.10	ttctaaccgctgtgtgttcagttgtactcgggtgtcatctg
1709	1749	0.87	taacgcgttttctgttgacaggactgtcaatcatggcctta
1732	1772	0.07	ctgtcaatcatggccttaaagatcgatacggccgaaacaaa
1880	1920	0.00	tttggttacaacacttaatagtcgccatttcgttattttgt
1910	1950	0.01	cgttattttgtaaatatgtagccatgattccaaggacgatg
1923	1963	0.00	atatgtagccatgattccaaggacgatggtgccaagaatga
2084	2124	0.63	atggcaacatccagctcttaggtatatctgtaataaagac
2141	2181	0.04	aaatattgaccatctgttcagcttttgcaatgaataatc
2244	2284	0.21	tgtgtgtgtgtgtgtacagtaaataaggaggtaaggagg
2285	2325	0.39	ggagtgtgtatttccctcagtggttttttgcgagacctcc
2299	2339	0.04	ccctcagtggttttttgcgagacctccgctgtccccgct
2334	2374	0.20	cccgtgcgcctccacaatagttgctggatttgtttccag

2354	2394	0.87	gttgcctggatttgttccagggtctagatctccagggacg
2369	2409	0.05	tccaggggtctagatctccagggacgtctctgaccggttga
2410	2450	0.98	cctctgctgcttccctctcagttccgatttgcagaccgctg
2423	2463	0.89	cttctcagttccgatttgcagaccgctgggacgtcgtgctg
2451	2491	0.09	ggacgtcgtgctgctcatcagcgggacagtgatggccatgg
2585	2625	0.07	gctggagtgagttttgctagagaaaatgatcaacaccaat
2640	2680	0.80	aacaacgccttttgccttcagattctactactactattctg
2665	2705	0.03	tactactactattctgaacagcacgttacaggaggacatgc
2675	2715	0.01	attctgaacagcacggttacaggaggacatgcaaaggtgtgt
2728	2768	0.00	ttcttactgtaactttaaagccctcctgaaggccctgaca
2739	2779	0.00	actttaaagccctcctgaaggccctgacattgcttatttt
2761	2801	0.49	ccctgacattgcttatttttagactttgaacctggcttctat
2847	2887	0.98	ctgcttattttatctatccaggcatacagttctgccagttt
2855	2895	0.04	ttatctatccaggcatacagttctgccagtttcctggtgt
2864	2904	0.03	ccaggcatacagttctgccagtttcctggtgttaacaggctc
2881	2921	0.95	ccagtttctggtgttaacaggctcttgggtggtctcacat
2918	2958	0.97	ccatggtgccgttttttgcagattcgccatttattactccg
2976	3016	0.14	ctgctggccgcctacatgcaggtgtccttctggaccataac
2998	3038	0.02	tgtccttctggaccataaacagccggcgccaggtgaaacgc
3054	3094	0.10	ttccactgcatcatgcagcaggagatcagctggtttgacgt
3138	3178	0.08	tcagagagttcccagcttcagcgttcacgctctgtacggct
3174	3214	0.01	cggctacgctcggaggtgtagatgatctgatggacgtgctt
3211	3251	0.03	gcttctttttccaatggcagcagcgtctacaagatccagg
3266	3306	0.00	aaggtggggctgctgatccaggcgtacaccaccttcacac
3379	3419	0.80	catttcggccgccttcttcagtaaagtaagaggctcctgct
3462	3502	0.00	gctgtgacctcttttgggaaggtgcttgcgtccttcaccag
3482	3522	0.02	ggtgcttgcgtccttcaccagtaaggagcagacggcgtacg
3545	3585	0.13	ggaagtgctgtccgcatcaggaccgtgttcgccttcagtg
3563	3603	0.69	caggaccgtgttcgccttcagtggtcagaccagagagattg
3575	3615	0.01	cgcttcagtggtcagaccagagagattgagaggtgcttac
3654	3694	0.00	agcagaagcatcacatttcagcttaaaaagctttttgttt
3690	3730	0.26	tgttttcctgatttcattagtaatgtgggttgactgaaag
3722	3762	0.00	gactgaaagacggtattgcagggggtgaatattgagaagcg
3827	3867	0.02	aacgacatgacatccatccagggacagtgatctgctagct
3845	3885	0.01	cagggacagtgatctgctagctttgctctccactcacacc

3892	3932	0.56	gttctgCGTGTGTCGCTTCAGATACCACAAGAACCCTGCGGG
4150	4190	0.15	cacttgTGTGTCCTTCATTAAGGTGTTTTTGTGCGTGCTTAT
4175	4215	0.00	TTTTGTGCGTGCTTATGGAGCGTTCAGCGTCGGACAGACC
4192	4232	0.00	GGAGCGTTCAGCGTCGGACAGACCTCTCCGAACATCCAGAA
4210	4250	0.02	CAGACCTCTCCGAACATCCAGAATTTGCCAGCGCCCGAGG
4313	4353	0.01	TATTATTCTCTTGAATTAGCATCTTTGCTGCTGGAAGCT
4331	4371	0.00	TAGCATCTTTGCTGCTGGAAGCTTGTGCTTCTGTGTTAGTG
4349	4389	0.18	AAGCTTGTGCTTCTGTGTTAGTGACTGCGCGCGTCAGCACC
4399	4439	0.17	TGGGAGTCCGCTTTGATGCAGAGAGGTGGGGGTTTTGCCT
4454	4494	0.97	CTGTTGGGTTCAATTCACAGGTCACTGGCTGATGTCACA
4477	4517	0.02	CACTGGCTGATGTCACACAGCTGGAGCCGGACTGATGTTT
4500	4540	0.01	GGAGCCGGACTGATGTTTCAGCTTCATGGCTCTAGCGTGGA
4514	4554	0.08	GTTTCAGCTTCATGGCTCTAGCGTGGACTGGCCTCATGTTT
4543	4583	0.78	GGCCTCATGTTTACCTCCCAGTGCTCAACAGGTGAATAGTA
4553	4593	0.74	TACCTCCCAGTGCTCAACAGGTGAATAGTAATGAACTGAA
4661	4701	0.00	ACGTGAGCTCCTAATCATTAGTCTTGCCCCCGTGCTGT
4711	4751	0.07	CTGTTGTGACCTCACTGTCAGCCGACAGGAAGTGTAGGTC
4780	4820	0.36	ATTGACCTTTTACTTTGTAAGGAAGAGGATTTGATTACCAT
4817	4857	0.15	CCATGGAAACCCCACTTCTAGCTCAGCGGCAGGATGTGGAG
4828	4868	0.02	CCACTTCTAGCTCAGCGGCAGGATGTGGAGGAATTCACGTG
4883	4923	0.00	TCTGACTGGATCTGGACGCAGAGGGTCTGGGCCTGTGTTA
4958	4998	0.97	CACATTTGTTTGTCTCCAGCTTCTAAGCTTGCGCGTCA
4967	5007	0.39	TTTGTCTCCAGCTTCTAAGCTTGCGCGTCACTTTGATGG
5005	5045	0.17	TGGTTTGCCTTGAATCTGCAGTTTCTTTCGAAGGGGATAAA
5017	5057	0.07	AATCTGCAGTTTCTTTCGAAGGGGATAAATATCACCTCAA
5275	5315	0.47	CCCCACTTCGTGTCTCAACAGAAGCCAAACATGACAGCTT
5300	5340	0.26	CAAACATTGACAGCTTTTCAGAGGACGGTTTCAAGCCGGAC
5330	5370	0.08	TCAAGCCGGACTTCATCAAAGGTGACATCGAGTTCAAGAAC
5373	5413	0.07	CCACTTCAATTACCCTTCGAGGCCTGAAGTCAAAGTGGGTG
5421	5461	0.81	TCTCCACGCATTGTTCCACAGACTGGGTCTGATGGCCGACT
5454	5494	0.81	GGCCGACTGTCTCATCTGCAGATCTTGAACAACATGTCTCT
5553	5593	0.01	ACCACCATCCAGCTGCTGCAGAGGTCTACGACCCCGAGGA
5602	5642	0.00	TGAGTCAGGGAGTCCATTCAAGTTCATTCAATTTAGGCACCT
5616	5656	0.32	CATTCAAGTTCATTCAATTTAGGCACCTCTGATAGCTACAGG
5629	5669	0.22	TCATTTAGGCACCTCTGATAGCTACAGGCTACAAAGTGGAA

5687	5727	0.54	ggaggcctttgctgctttcagaacattctggtaaatattca
5751	5791	0.95	tggattcactctcctgtgtaggtatztatcgacgggtcacga
5953	5993	0.11	acttcatcatgaacctccagacgtgaggcctcatttatgg
6039	6079	0.01	tacattatgatctctgagcagaagtttgagacgctgggg
6195	6235	0.01	gagagcgagaccatcgtacaggctgctctggacaaggtaca
6245	6285	0.97	atgctccccctctccatgtaggaccagcctttggatctca
6268	6308	0.15	cccagcctttggatctcaaagatatttctcagtatgggcag
6279	6319	0.00	gatctcaaagatatttctcagtatgggcagtatgtgtttag
6371	6411	0.00	ctggtgaccaaagtccaacaggacgtatttatgtaactgtc
6401	6441	0.04	atgtaactgtcccagtcataagaatgaaatgcttttgaagat
6419	6459	0.00	tagaatgaaatgcttttgaagatgtcaaaggtctacacctc
6428	6468	0.00	atgcttttgaagatgtcaaaggtctacacctctggataatg
6466	6506	0.00	atgaatacaaccatttcaaaggagcaatgctgttttattac
6514	6554	0.01	ttattgaccatattatcaaaggtcaacattgtcacgctagc
6593	6633	0.75	gttggtttttgtaaactccaggtccgactgggtcgcaccac
6646	6686	0.04	tcaccgactctcgaccatcagaaacgccgacatcattgctg
6673	6713	0.04	cgacatcattgctggattcagtaatggcaaaatcgtggagc
6755	6795	0.10	catggcctggtgaccatgcaggtgaccccccccccaacg
6846	6886	0.92	gaaattctctccttgcttcagacgttcacacaatgtggagga
6933	6973	0.00	gtggaaaagaccgtctcccagtcgtccatcatcaggaggaa
6947	6987	0.56	ctcccagtcgtccatcatcaggaggaagtcaccagaggggt
6985	7025	0.30	ggtcctcgtttgccgcgtcagaaggaaccaagaggaaaag
7103	7143	0.72	ctttacctgatcaccgtccaggagtccagacgctgccgtct
7111	7151	0.00	gatcaccgtccaggagtccagacgctgccgtctctttctct
7187	7227	0.50	aaagttgattttgattgtagattcctaaatagcagacagg
7202	7242	0.02	tgttagattcctaaatagcagacaggtgctacactcaggtg
7219	7259	0.01	gcagacaggtgctacactcaggtgataacatcgatggttat
7258	7298	0.76	attacacatcctgtgttccagattacagtccaatttacacc
7312	7352	0.05	gctcttcattctgacaattagatttgaaaaatgataacaat
7398	7438	0.35	ctccgtgcgctggttccgcaggatgaagacgttcccgcgct
7432	7472	0.18	ccgacgtgtcgttctttaaagtgtgcatctgaacatcccc
7536	7576	0.59	ttcgccatcctcttctccaagatcatcactgtaggtccaac
7549	7589	0.74	tctccaagatcatcactgtaggtccaacgcgctgaccgtgg
7609	7649	0.42	caaatgttgcgcttctgcagaaggcttgaaaaaaaagcac
7649	7689	0.99	cgttttgcttttttattgcaggtgtttgcggatccagaccg

7665	7705	0.01	tgcaggtgtttgcggatccagaccgtgattctgtcaggagg
7681	7721	0.12	tccagaccgtgattctgtcaggaggaagagtgaattcattt
7754	7794	0.97	tttgtcaccatgtttttacaggtaagtgtgttttgacctg
7832	7872	0.03	tggccgaaaaactttcaacagggttactgtttcggtaaatc
7977	8017	0.83	ctacggctacgtgcccttcaggccccgccacaaatacagt
8054	8094	0.58	tcaatcctctctgtgttgaaggacctcagctggtacgacaa
8107	8147	0.38	cgttggcgctctcaccactaggctggctgccgacgccgcc
8173	8213	0.01	ggacgtgacatttgacctcagcgaggcctaattaacggagg
8211	8251	0.93	aggatgctcctctctgctcaggctgcaggggtgcgcctggc
8244	8284	0.00	cgctggcgacaataatgcagaacttcgccaacctgggcac
8451	8491	0.85	gttaaattcgaccttttacagggtgccattgtggcataag
8501	8541	0.98	aagtgtggtttctccctgcagatcgccacagaggccatcga
8511	8551	0.05	tctccctgcagatcgccacagaggccatcgagaacgtcaga
8548	8588	0.12	cagaaccgtcggtccctcagcagagaaccaaatttgagt
8697	8737	0.51	ttctcttgattttgcgacaggaactcccagaaaaaggccc
8749	8789	0.95	ttaacctactccttctcccaggccatgatctactttgctta
8792	8832	0.01	ctgcctgtttccgcttcggagcctggctgattgaagcagga
8848	8888	0.06	ggagtgttccctgtgagtacagacacagcacggcctgatgat
8886	8926	0.05	gatggcttctgtatggtgaagggtgacatttacactgagcac
8989	9029	0.14	tggcttctaatacattgaccagggtcaggacagcatcacgct
9078	9118	0.91	cttcacattctttcaatttagaaaggatataaaagtaacat
9139	9179	0.07	tacctctcacacgcatacagatttcccttttagactttagc
9151	9191	0.91	cgatacagatttcccttttagactttagcatcttgatattt
9262	9302	0.01	aacgcaccaaattcaatctagggtggaagaactgattaaa
9335	9375	0.00	tttaattaataacttctgaagtcttcaaaaataagcactg
9377	9417	0.00	acaacatattttatcacttagtcattaaaggctttcccaga
9386	9426	0.27	ttatcacttagtcattaaaggctttcccagacgatatgga
9396	9436	0.01	agtcattaaaggctttcccagacgatatggaaccgtttatt
9440	9480	0.14	ttattattcaaacatccacaggcaaccgttgaagaatgcct
9479	9519	0.17	ctttgtttcatgtttgcttagcgttatgcaatagggtttga
9492	9532	0.12	ttgcttagcgttatgcaatagggtttgataatgttacgaaa
9545	9585	0.94	ttatttaactttgcttcccagggtggtttctgcggtgctgt
9685	9725	0.08	ccgccattgataacctctcagaggaggggacgtctccggta
9761	9801	0.00	gttccacacgcacgctaacaggcggtcttctcgcccacct
9785	9825	0.79	gctcttctcgcccacctccaggaaaaatacgcaggtaatgt

9863	9903	0.07	cccgatgtgaccataactccaggggctgaacctgaaggtgaa
9916	9956	0.00	tctggccttggtgggcagcagcggttgtggaaagagcacca
9970	10010	0.04	ggagaggttttatgaccccagagaggggagagtggtgagaa
10067	10107	0.29	tgcctcttcacactactgttagttcgataatgattagctcgt
10111	10151	0.93	tgttttattgctgctctccagcgttgacagttataggacaga
10119	10159	0.73	tgctgctctccagcgttgacagttataggacagagatcatcg
10151	10191	0.00	agatcatcgagatctgtgcagtacggatgtttgttctttac
10182	10222	0.19	tgttctttactgcgctacagtctgtttatggtgatcagta
10200	10240	0.00	cagtctgtttatggtgatcagtaacaagattcatagttatc
10241	10281	0.77	caaaatctttacactcacagaaactatatacagacccac
10280	10320	0.01	accactgagctgctttgtcagaacattcttgtaacgtcatt
10308	10348	0.70	ttgtaacgtcattatctccagatcactaactcataactgg
10348	10388	0.58	tggtctgttgactctctgcagatttgcatcagacaaatg
10361	10401	0.56	ctctgcagatttgcatcagacaaatgtagctgcagggat
10377	10417	0.01	tcagacaaatgtagctgcagggatcctctttcgcaactgt
10399	10439	0.08	gatcctctttcgcaactgtagaagcttagaacaggactaaa
10473	10513	0.02	tatgtaaactctgattaacagatttattatatgctcattgg
10496	10536	0.79	ttattatatgctcattggcaggttgagaaaattgatcacia
10535	10575	0.08	aagccaaatttgccggttacagatgatttttttattgaattt
10560	10600	0.83	tttttttattgaattttgcagcgtggcgccccctgctggt
10593	10633	0.04	ctgctgtttatacatatagacaaagcatttaaaacattc
10653	10693	0.00	tcattcattttcattacaaaaggacattttaaaggcttctaca
10699	10739	0.03	acaaaactattttgatgcagtatgactgacatcactgcta
10731	10771	0.10	tcactgctacctctgatctagtttagctcccataaaatggt
10763	10803	0.13	taaaatgttttggtttttaagacaccatacaaagctaaagt
10801	10841	0.98	agttctcctttccctttcagtcactggacgggtgtcaacgt
10855	10895	0.09	attcactggctgaggtctcagatcggcatcgtctcccagga
10873	10913	0.43	cagatcggcatcgtctcccaggagccgggtgctgttcgactg
10998	11038	0.04	acatccacagcttcatcgaagggtgcctcaggtagcggct
11009	11049	0.36	ttcatcgaagggtgcctcaggtagcggctgtgaatcaggg
11111	11151	0.10	atacactgtctctgagcagagatacgacactcaggctgg
11181	11221	0.00	agaagcagcgtgtcgccatagcccagccatcatccgcaac
11279	11319	0.21	gtgagactttattcacactagtgtacgtagtccataagggc
11289	11329	0.03	attcacactagtgtacgtagtccataagggcctgcacaa
11297	11337	0.00	tagtgctacgtagtccataagggcctgcacaaactgcac

11330	11370	0.98	cactgcacctcttgtttccaggtggtgcaggaggcgttggga
11339	11379	0.03	tcttgtttccaggtggtgcaggaggcgttggaccaggccag
11385	11425	0.00	gcaggacgtgcatcgtggtagcccaccgtctgtccaccatc
11408	11448	0.79	caccgtctgtccaccatccagaacgccgactgcatcgctgt
11435	11475	0.49	gactgcatcgctgtgttccagggaggagtggtggtggaaaa
11516	11556	0.01	cacatgctggtcaccaaacagatgggctatcacaacgactg
11565	11605	0.34	cccgtgttttccagctttagaagaacatttagagtctga
11630	11670	1.00	ctcgtttctcttgctcttcaggctgagacatccgtccttct
11654	11694	0.02	gagacatccgtccttctcaagcttgaacattcgagcactt
11799	11839	0.43	cgctgcggtgcctgttaacaggctgtttactgcaattatgt
11826	11866	0.28	tactgcaattatgtctttagttagttagttagttcaattcttt
11835	11875	0.02	tatgtctttagttagttagttagttcaattcttttatattgatc
11873	11913	0.00	atcatatatttgtggttgaagtctgaactgaaacactgaaa
11911	11951	0.00	aaaaatgtatcctgtaaccagactggcaaaagacttcagtt
11939	11979	0.07	aaagacttcagttcaattcagagtgaggacgtgttgacagc
12006	12046	0.16	attgttggttggtttttaaagagaatgtgggggtgtagaca
12044	12084	0.25	acaggacgcctcttccccagcgggggtcttatgtgcttt
12131	12171	0.04	caaagcatcgatccacgacaggaccagtaagatggagcgaa
12166	12206	0.00	agcgaacgtgccaccacacagcagaaaagacaccaagacaaa
12257	12297	0.01	aaaataacagccttaatgcagctctaatgtacagtaaatac
12346	12386	0.03	tcagtctaacatttatcacagtcggaggcttgagcttcac
12491	12531	0.01	gggaaggttgcatcgatcagaggtttgttggtgtagggag
12539	12579	0.10	gagctgagacgttgttgccaggtgcagcatctcgtgcagac
12557	12597	0.09	caggtgcagcatctcgtgcagacagctgctgaagacgttga
12606	12646	0.01	acggcgtcggctctggcggcaggacttcaggccgatatgga
12615	12655	0.01	gtctggcggcaggacttcaggccgatatggaaatcacaat
12715	12755	0.11	agcagagtgatccgtcagtaggcccactgtcctctcggat
12742	12782	0.02	ctgtcctctcggatcaggtagaagaagccgtagccgtgggg
12786	12826	0.01	catcggcgccaccgcgccagaactgtggtgtagcccacca
12807	12847	0.03	aactgtggtgtagcccaccaggctggacgacagcacaaggt
12854	12894	0.07	caccgccactgcacaacacagacagcatggaaaaaccagcc
12890	12930	0.00	cagcccgtttctgaataaaaaggtgagtgaagctaactagcc
12944	12984	0.04	ctagcagcgtttccacgcagtacctcttggcatagagcgg
12959	12999	0.43	acgcagtacctcttggcatagagcggatctaagaagagttc
12987	13027	0.00	ctaagaagagttccggaacaggacgtccctcctctttggcg

13018	13058	0.00	ctctttggcgatgagatacaggccgagaaggtgcctgtcga
13049	13089	0.05	tgcctgtcgaagcctgcacaggcgagagcgacgacgtaaca
13100	13140	0.96	ggacgccgttctccccattagactcagacaaagtgaagggc
13157	13197	0.07	ttcctgggcctcggccatcaggttgttgtgtttctcaaagg
13176	13216	0.16	aggttgttgtgtttctcaaaggccagctgcatggctttcct
13201	13241	0.83	gctgcatggctttcctcttagcgttgcctgaaacagaaat
13217	13257	0.12	cttagcgttgtcctgaaacagaaatcgccacgcgtcaagcc
13262	13302	0.01	cctcctgccacccacggcagcgtctcaacacacctcgac
13330	13370	0.06	ttcaccgcctccacggtgcagggcctcatggtctccgtcct
13360	13400	0.11	gtctccgtcctgccgtgtagaacttgcgagtcaggtgt
13387	13427	0.09	cgagtcaggtgtctcgtagcaacacctggcctgagagg
13491	13531	0.02	tacctctggtgcagtttaaagtacgccagctgcatcgccag
13511	13551	0.01	gtacgccagctgcatcgccagttggatgaacgtgtccgggt
13535	13575	0.00	gatgaacgtgtccgggtgcagctttttctgcttgatggcgg
13567	13607	0.01	tgatggcggcttttccgaaagccgtgaaggcgtaacagaca
13595	13635	0.02	ggcgtaacagacaacctgcaggtcctgcgcctggacaacag
13615	13655	0.01	ggtcctgcgcctggacaacagacggctgtcactccaccggc
13735	13775	0.84	tgcggacttttctgctccacagtaaacaccagctcctcgggga
13745	13785	0.00	ttcgtccacagtaaacaccagctcctcgggaggcggcagga
13763	13803	0.00	cagctcctcgggaggcggcaggactctgactgtgtccacac
13789	13829	0.13	tgactgtgtccacacccttagaaatatacaaatgttaatag
13829	13869	0.10	gcgcacgccttcaatttaaagagcggcagagaatgggaatgg
13902	13942	0.00	attcgtggtccacgtaccagcacatggtcaccagcaccat

1. 比较结果二和结果三发现与人 MDR1 同源的 cosmid 124A22 中的大部分序列的起止点均符合真核生物的“gt.....ag”法则，只有少数出现 1-2bp 的差异，这样得到了人的外显子 6、7、8、9、10、11、12、13、14、15、17、18、19、20、21、22、23、24、25、26、27、28 对应的河豚鱼序列。
2. 根据结果二和三，以及严格的“gt.....ag”法则，推算出外显子 16 可能对应的 cosmid 124A22 的序列。
3. 将每段序列剩余的、与人 MDR1 无同源性的序列单独列出，降低搜索严格度，继续以 blastx 软件来分析，找到了人 MDR1 外显子 4 对应的序列区，在 exon4 和 exon6 之间的序列中，根据“gt.....ag”法则和内含子类型的匹配来估算 exon5 对应的同源序列位置。

4. 同源性搜索并未发现人 MDR1 exon1、exon2 和 exon3 的河豚鱼同源序列。

四、通过启动子分析软件（网址：http://www_hgc.lbl.gov/projects/promotor.html）来预测 contig2 中所有可能存在的启动子序列，结果如下：

Start	End	Score	Promoter Sequence
236	286	0.90	CTACGTAGTCCATAAGGCCGCTGCACAACACTGCACCTCTTGTGTCCAGG
478	528	0.93	AACACAGTGGATAAAAAGACGGTACAACACTATGCCAATATCATTTTTACAT
1965	2015	0.89	TGTCAGGGTCATTA AAAATGATCCACAGGAACATCTTTGAAGCGCTTCTG
2273	2323	0.93	TTTATATGTATATATGTACACAGTTATGGCTCCTGTAGTGACAACAAAA
2974	3024	0.83	GGGAGCAATTGTTAAATCAGGGCAGATATTTTTCAATAAAACCTCACAG
3396	3446	0.87	ATTGCTAGTCGATAAAGGCTCGTGCCTCCTTACCTAGCCGAGTTTGGTT
4711	4761	0.97	AGGAGCGGCGTATAAAGTCTACAACATTATCGACCAGTGAGTCCAGTGG
6625	6675	0.89	TCCGCAACCCATAAAATCCTCCTGCTGGACGAGGCCAGTCTGCCCTCGAT
7058	7108	0.85	AAAAAAACCATTAAAAATGGGTTTTCTCTACAGGTCGACTGGGTCCAC
7543	7593	0.80	GCTGTGGTCCAGAGACTTTACAAAACAGGCGGCTCGCCTCAGATTCAGAC
7552	7602	0.97	CAGAGACTTTACAAAACAGGCGGCTCGCCTCAGATTCAGACTTAACAGTG
7806	7856	0.81	TGAGTGGCGTATATTTTGGTAGGAACCATCTGCGCCATGATCAACGGTG
8135	8185	0.85	TTTGAAAGGAAAAAAGCATAGAAAAAGCCCGGGTTGTCCAACAATAGCA
8144	8194	1.00	AAAAAAGCATAGAAAAAGCCCGGGTTGTCCAACAATAGCAGCGGTTTAG
8577	8627	0.88	ATCGCTGTGGTCATAAAGGGGTGAAATGATCATTAAACGACTCAAGTAG
8980	9030	0.96	GCAGCATGTAAAAATAAAGCAGCCCAACATTTAAATGGCAATTTAATG
9356	9406	0.91	GATATATTTGTAATATGTATCTCGGGCACTGGACTTTAAATAGATAAT
9414	9464	0.86	TTGGATGAATAATAAGGTGCCGACCCAAGACACGGCTTTTATTTCCACAC
10712	10762	0.95	ATTCCCTTCTGTATATAATTTCCOCTCTGTATATAATGTATAATGTACA
10794	10844	0.90	TTTAGTGTGTATATTACTGGCTACTTCTGTAACAGTGAATTTCCCAA
10967	11017	0.98	CTTAATCTATTA AAAATCACCACAAATCTTAATGTTGTTGTGCACACCG
11427	11477	0.85	GTCGATGCTTTAAAAATGACAGCAAGCATGCTAACACACGGTTCTTCTCTG
11573	11623	0.81	TCTGGAGGTGCAAAAGGAGAACTCTGGCCTTGGTGGCAGCAGCGGTT
12615	12665	0.90	CTACGTAGTCCATAAGGCCGCTGCACAACACTGCACCTCTTGTGTCCAGG
13085	13135	0.97	ATGAAAAATAAAAATAAGCCCTCCTACTGAATCTTCTTGAGTTTCTTTTC
13211	13261	0.84	ACTAAAAATATATATATATATATATATACTGGAAGTGAAGTGAAGCA
13220	13270	1.00	ATATATATATATATATACTGGAAGTGAAGTGAAGCAATAAACCCCT
13319	13369	0.98	AAAGTAAATCTATATGTGAGGGAGGGAGAGCAGGAAATAGAAGAATTTA
13362	13412	0.88	GAATTTAGTGTAAAAGTGTGTTAGCGTGAGAATCTAGCGGTTTGATCG
13498	13548	0.86	GCCCTCTGTCCAAAATAGACCCAGCTCGCAGGGCTCTGGGACACCGGCAG

13697	13747	0.85	TTTTAATTTAAATAAOCGGTTTTCTGTTGACAGGACTGTCAATCATGGCC
13910	13960	0.94	CGTTATTTTGTAATATGTAGCCATGATTCCAAGGACGATGGTGCCAAGA
14049	14099	0.97	TAACATGCTCATAAAAACGGCCAATATTAATCAACAATGGCAACATCCAGT
14255	14305	0.86	GTGTGTACAGTAAATGAGGGTAAAGGAGGGGAGTGTGTATTTCCCTCA
14540	14590	0.97	CGGACAGTTTTATATACGCTGACATGGCCCAACACAACGCAAGTGGCTGG
14728	14778	0.91	TTCTTACTGTAACTTTAAAAGCCCTCCTGAAGGCCCTGACATTGCTTATT
15003	15053	0.95	TTCTGGACCATAACAGCCGGGGGCCAGGTGAAACGCATCCGAGCTTGTT
15638	15688	0.97	GCTACATTTTATAAAGAGCAGAAGCATCACATTTTCAGCTTAAAAAGCTTT
16103	16153	0.96	ACTTTCCGCATAAATGAGGGTGGTGAGGGTCTGACAAGCCTGCGTGCAC
16248	16298	0.91	AGGAGCCGCCTATAAAGTGTACAGCATCATCGATAACGTAAGGAGACCAC
17032	17082	0.90	TCGAAGGGGATAAATATCACCCCTCAAACATGAAACAAATTGCTAACAGCC
17206	17256	0.89	ACTGGATGTTATAAGAATCGGCCCGCTCTCGGCTTTGTTTCTCGG
17248	17298	0.96	TTTCTCGGGTTAAAAATATTGCACGTTCCCACTTGTGTCTCAACAGAA
17671	17721	0.99	GCOGTGACCATAAAAAGGAGGCCTTTGCTGCTTTCAGAACATTTCTGGTAA
17765	17815	0.80	TGTGTAGGTATTTATCGACGGTCAAGACATCCGCTCCCTCAACATCCGCT
18142	18192	0.89	TCCGCAACCTAAAAATCCTCTGCTGGACGAGGCCACGCTGCCCCTGAT
18741	18791	0.81	TAAAGGGAGTCTATCATGGCCTGGTGACCATGCAGGTGACCCCCCCCCC
19436	19486	0.90	CGTGTGTTCTTTAAAGTGTGCATCTGAACATCCCGAGTGGCCCTACA
19628	19678	0.90	AGAAGGCTTGAAAAAAGCAGCTTTTGCTTTTTTATTGCAGGTGTTTGC
20580	20630	0.89	AATTTGAGTGTATATGAGGAGAATCTCAGAGTCCGTACAAGTATGAC
20716	20766	0.95	AGGAACTCCAGAAAAAGGCCACGTTGTAACGGCTTAACTACTCCTCTCTC
21892	21942	0.89	CCTGAAGGTGAAAAAGGGAGAACTCTGGCCTTGGTGGGCAGCAGCGGTT
22044	22094	0.95	CCGTTATCTCCATATAACCCCGTGCCTCTTACCTACTGTTAGTTGAT
22259	22309	0.83	CAGAAACACTATATCAGACCCACTGAGCTGCTTTGTCAGAACATTCT
22624	22674	0.88	TAAAACATTCTTAAAAAAGCATTTAGTTTCATCATTTTCATTACAAAAG
23302	23352	0.90	CTACGTAGTCCATAAGGCCGCTGCACAACACTGCACCTCTTGTTTCCAGG
23686	23736	0.94	CGAGCACTTCTATAAAACCAACCTTTACTGCTGCTGAOCTTTAACCTCAC
23781	23831	0.80	AACGCTGGATTATTTATGCGCTGCGGTGCCTGTTAACAGGCTGTTTACTG
24893	24943	0.95	CCCGTTTCTGAATAAAAGGTGAGTGAAGCTAACTAGCCTTAAGACGGAAC
25834	25884	0.99	CGCCTTCAATTTAAAGAGCGCGAGAATGGGAATGGTGACTIONTCCAT
26142	26192	0.97	TGAGCTGCGTTCAGAAAAAGCAGGGGCGCACAGTCAGACACCGTTTAGG
26698	26748	0.83	GAAATGGAAGTAAATAGGGGTGTCACAACCCCTGACTGCCTCCTAGTGTT
26964	27014	0.87	GTACTTACTGTATAATTACTACCTGAACAGTTCTGGGGGCGTCAGTATTT
27845	27895	0.90	TAAATAGCTTTTATAAAGCCTGACACATCAGAAAATAGCCAGAAAATTCT
28280	28330	0.88	GTTAAGTGATCTATAAGTGTGATACTTCATTATTCACTCCTGGCAGTGA

28366 28416 0.88 CAGTTTAGTTTAAAAATATGCAGATTTAATTAACAAATAATAACACAGAA

Predictions for the reverse strand of seq0 :

Start	End	Score	Promoter Sequence
28170	28120	0.95	ACACCACCCCGTAAAAGAGGCTGCAGTGTGCTCTTGTGCTAAAAAGTGC
27869	27819	0.93	TGTCAGGCTTTATAAAAGCTATTTATTTTCAGGCTCGGTGATGGTAAAT
27190	27140	0.97	TAGCTACTTTAAAAACAGGTGCGAGAAAACGGCACTTTCGACACGCGTATAT
24454	24404	0.81	TTTTGGGCTGTAATGTGCTCCTTGATTATGTAGAAAACGATGCCGAATG
24088	24038	0.98	CGTTAAAGCACATAAGACCCCCCGCTGGGGGAAGAGGCGTCTGTCTACA
24033	23983	0.82	ACATTCTCTTTTAAAAACAACCAACAATCAGAAATCAATCAAGTTTCATA
23805	23755	0.98	CGCAGCGCATAAATAATCCAGCGTCTGCCAACTTCAGTCACTGCAOCTG
23124	23074	0.92	GAGACAGTATGTATATGGTGTAAATGGTCATGGGCAGCTCCATGGGAATC
22619	22569	1.00	CTTTGTCTATATATGTATAAACAGCAGGGGGGCCACGCTGCAAAATTC
22609	22559	0.91	ATATGTATAAACAGCAGGGGGCGCCACGCTGCAAAATTCATAAAAAA
22154	22104	0.86	ATCTCTGTCTATAACTGCAACGCTGGAGAGCAGCAATAAACACATACG
22048	21998	0.89	AACGGGAGGAAAAAAAACACAGCACATTGTGTTTGTCTTCTCACCCTC
21991	21941	0.86	TCTGGGGTCATAAAACCTCTCCAGCAGCTGGATGGTGGTGTCTTTCCAC
21704	21654	0.96	TGAGAGGTTATCAATGGGGGCTTCTGTATTATTAGCATCATCAGGTAGG
20605	20555	0.81	ATTCTCTCATATAAACTCAAATTTGGTTCTCTGCTGAGGGACACGA
20024	19974	0.97	AATCAAACTGTATTTGTGGGGGGCCGAAGGGCAGTAGCCGTAGCA
19781	19731	0.97	CACCTACCTGTA AAAACATGGTGACAAATGACACACAGCCAATAACGACA
19167	19117	0.96	AACGAGGTGTAAAAGAGAGAAAGAGACGGCAGCGTCTGGACTCCTGGAC
18412	18362	0.90	GGACAGTTACATAAATACGTCCTGTTGGACTTTGGTCAACAGGAGCTCAC
18050	18000	0.97	GATCATAATGTAAATATTAGCGTTAGAGCGGACGGAGCCAGTGATAAAT
18002	17952	0.94	ATATGGAAGCCATAAATGAGGCTCAGCTCTGGAAGGTTTCATGATGAAGT
16964	16914	0.96	AAATGTGAAATAAACAGTGCGGTAATGTTACACACAGGOGTAAACACAG
16273	16223	0.88	TGCTGTACACTTTATAGCGGCTCCTCGGGGCTGGCAAAATTCGGATG
15659	15609	0.94	TCTGCTCTTATATAAATGTAGCTTTCACGTGCAAAGACGTTTATGTAAGC
15451	15401	0.83	GGGAACGGGTACAAAGAGAGCAGAGGGTTAAGCAGGAGCCTCTTACTT
15394	15344	0.88	GAAGGCGGCCGAAATGGCCAGCGGGGCTCAGGCCAGGATGACCAGCG
14961	14911	0.86	AGACGGAGTAATAAATGGGAATCTGCAAAAAACGGCAACATGGTGAGAC
14945	14895	0.95	GGCGAATCTGCAAAAAACGGCAACATGGTGAGACCACAAAGACCTGTTA
14795	14745	0.86	GCCAGGTTCAAAGTCTAAAATAAGCAATGTCAGGGCCTTCAGGAGGGCTT
14563	14513	0.99	TGTCAGCGTATATAAACTGTCCGTCATTTCTCCAAAGACAATGCACATG

12123	12073	0.96	AAATGTGATGTATAAAATATGCAATGGTACGAGTAAACGCTCCACTTTGC
11672	11622	0.86	TCTGGGGTCATAAAACCTCTCCAGCAGCTGGATGGTGGTCTCTTTCCAC
11307	11257	0.84	AGTTGGGAGCATAAGTGTGGCCTCGCCGACAGCCATGGGCGGTACAGC
10756	10706	0.85	ATTATACATTATATACAGAGGGGAAATTATATACAGAAGGGAATATTAG
10114	10064	0.87	ATTTTCTCGTATAAAGCCACAAATGTCGGTTCTCTGGTGAGGGACACCA
9514	9464	0.96	CAAAACTGGTTTTAAATTAGGCTAGGTTGCTGATCCGGGCATGATCTGCA
9466	9416	0.99	CATGTGTGGGAATAAAAGCCGTGCTTGGGTGGCACCTTATTATTTCATC
9099	9049	0.81	TAAAAAGCAATTTAAATGCTGCGATCTTATTCTCTCTGATGAAGGAGTGG
8281	8231	0.85	GTGAAGCCTCTAGAAATGTGCTGACAGGACTGACAATAGCATAGAATCCA
7640	7590	0.98	TAACCAGACCTATAAAGGTCAGCTGATCCTGGCTTATCTCACTGTTAAGT
6533	6483	0.97	GATCATAATGTAAATATTAGCGTTAGAGCGGGACGGAGCCAGTGATAAAT
6485	6435	0.94	ATATGGAAGCCATAAATGAGGCCTCACGTCTGGAAGGTTTCATGATGAAGT
6081	6031	0.97	TTCACCTCTCTCTATATAGTGCACATGTGCACCTACTGTCAATGGTTGAA
5817	5767	0.85	TGTCGTTTAAGATCTGCGGATGAGAGAGGTGAGTCGGGCAGGGACATGG
4736	4686	0.84	TGTTGTAGACTTTATACGCGCTCCTCTGGCACTGGCAAAGGACTGGACG
4098	4048	0.97	TATAGGAGTATAATAAGAGGGGAGTAATCGTGTCTATGGATGGGGGTGCC
3835	3785	0.83	CAAAGGAAGTGATAAAGGTGGAGGCCCTGGATCAGCAGACCCGCCTTG
3481	3431	0.90	TGGAGAATCTAAAAAATAGAGGTTGGAGATTAGAGTAACCAACTCGGCT
3087	3037	0.85	AGTCAAAAGGGTTATAAAGTACGAAGAAATGAACCATGACATTGGCTCCT
2189	2139	0.98	AACACATCAATATATGAATCGAGGGGAAAAGACCTACCTCAAATTCTGTT
2067	2017	0.95	ATTTATTTTTTAAAAAGTAGCGTTTCCAAGTAAAAATCCCAGGGATCTTA
2047	1997	0.94	CGTTTCCAAGTAAAAATCCCAGGATCTTAAGACAGAAGCGCTTCAAAGA
1643	1593	0.80	AGGTTGTTCTAAAAACCCTGCTGGACCCAGATTCTGTTCACTCTAACG
58	8	0.91	GAGACAGTATGTATATGGTGTAATGGTCATGGGCAGCTCCATGGAAACG

- 根据前面获得的 exon4—exon28 的对应序列位置, 推测出 1bp-14kb 和 12kb-26kb 可能的启动子位置:
 1bp-14kb 的启动子位置: 1965bp—2015bp
 TGTCAGGGTCATTA AAAAATGATCCACAGGAACATCTTTGAAGCGCTTCTG
 12kb-26kb 的启动子位置: 13220bp—13270bp
 ATATATATATATATATACTGGAACTAGTGAAGCAATAAAACCCT
- 根据启动子的位置和 exon28 对应序列的位置, 找到可能存在的起始密码子和终止密码子:

	起始密码子	位置(bp)	终止密码子	位置(bp)
1bp-14kb	ATG	2080	TGA	12868
	ATG	2233	TGA	12868
	ATG	2278	TGA	12868
12kb-26kb	ATG	1332	TGA	11555
	ATG	1443	TGA	11555

五、利用 Grail 软件(网址: <http://compbio.oml.gov/Grail-1.3/>)来分析 cosmid 124A22 中所有可能存在的加尾信号:

Str	Start	End	Score
f	980	985	0.90 (fuguMDR3)
f	1251	1256	0.80
f	13091	13096	0.88 (fuguMDR1)
f	19384	19389	0.65
f	29174	29179	0.90
f	33425	33430	0.80
f	38368	38373	0.65
f	38773	38778	0.65
r	6465	6470	0.88
r	8405	8410	0.80
r	8880	8885	0.90
r	9069	9074	0.80
r	10178	10183	0.80
r	12032	12037	0.65
r	17383	17388	0.80
r	17452	17457	0.90
r	26121	26126	0.90
r	29532	29537	0.80
r	30496	30501	0.80
r	35650	35655	0.80

Note: Str=strand; f=forward; r=reverse

六、河豚鱼的两个完整的MDR1 cDNA: fuguMDR1 cDNA 和 fuguMDR2 cDNA 分别对应于序列 1bp-14kb 和 12kb-26kb, 它们的 exon 位置如下:

外显子 序 号	fuguMDR1		fuguMDR2	
	位置	大小 bp	位置	大小 bp
4	2616-2781	166	2431-2590	160
5	2911-2959	49	2661-2709	49
6	3474-3669	196	2939-3130	192
7	3772-3924	153	3232-3403	172
8	4133-4257	125	3483-3607	125
9	4377-4548	172	3913-4084	172
10	4634-4747	114	4171-4284	114
11	5608-5718	111	5296-5406	111
12	5805-5930	126	5475-5600	126
13	6255-6458	204	5772-5975	204
14	6543-6713	171	6060-6230	171
15	7091-7252	162	6614-6775	162
16	7344-7538	195	6867-7034	168
17	7753-7899	147	7419-7565	147
18	7979-8083	105	7670-7774	105
19	9203-9295	93	7853-7930	78
20	9569-9652	84	8075-8158	84
21	9726-9929	204	8232-8435	204
22	10031-10131	101	8522-8622	101
23	10229-10369	141	8718-8858	141
24	11243-11399	157	9566-9722	157
25	11487-11684	198	9806-10003	198
26	12136-12342	207	10822-11029	207
27	12445-12591	147	11132-11278	147
28	12664-12867	204	11351-11554	204

它们的 cDNA 序列和其对应的氨基酸序列为:

1bp-14kb 的预测 cDNA 序列

TTCCGGTTTCAGATGGTTGGGACATCTTGATGGTCACCATCGGAGTGCTGATGGOCATTGTGAACGGGCTGGTGAATCCTC
TGATGTGTATOGTGTGGTGAGATGACTGACAGCTTCATCCAGGAAGCCAACTGTCCAAAACCACAACAAGCAACCC
CACTTCAACAATGTCTAATGTTTTAAGAGTGTGCTGCATCATTOGCCATCTATTCTCCATCTACTACTCCATCTTGGGGTTT
GCTGTGCTGGTAGTGGCGTACCTGCAGATGTCTCTGTGGACCTAACGGCCGCGGGCAGGCCAAAAGCAATTCGGAGTTGT
TTTTCCAGGCATCATGCAGCAGGACATCAGCTGGTATGACGTGACTGAGACAGGAGCTCAACACGGCTCTCACAGAGTG
GGAGGGCATOGGTGACAAGGGGGTCTGCTGATCCAGGGCGCTCCACCTTTATCACTTCTTTGTTATTGGTTTTGTACAT
GGATGGAAGCTCACCTGGTCACTCTGGCCATCAGCCCTGTGTTGGGTCTCTCAGCTGCCCTTTACAGTAAGTTGCTGACAA
GCTTCAACAGTAAGAGCAGACAGCGTACGCCAAAGCTGGAGCTGTGGCAGCGGAGGTGCTATCCTCCATCAGGACTGTGTT
TGCTTCAGTGGCCAAAGAAAAGCCATCAAAGATATCATAAGAACCCTGGAGGATGCGAGGGACATGGGAATAAGAAGGGA
GTTGCTGCTAACACGGCCACAGGCTTCTCCTTTCTGATGATCTACCTGTCTATGCTCTGGCCTTCTGGTACGGGACTACTC
TGGTCTCAACAAAGAGTACACCATTTGGAATTTACTGACTGTGTTCTTTOGTOCTCTAOGGGGCATACATTATTGGACA
GGCCTCTCCCAACGTCCAGTCTTTGCCAGTGCCAGAGGAGCGGATATAAAGTCTACAACATTATOGACCACAAACCGAAT
ATTGACAGCTTTTCAGAGGACGGATACAAGCCTGAATACATCAAAGGTGACATTGTATTCCAGAACATCCACTTCAGCTACC
CTTCGAGGCCAGAAATTAATCTTAAACGACATGTCTGTTTCATGTGAGGAACGGACAGACCATTGCTTTGGTGGGAGCAG
CGGCTGTGGTAAAGTACCACCATCCAGCTGCTGCAGAGGTTCTACGACCCCCAGAAAGGATCCATATTTATOGAOGGTAC
GACATCCGCTCCCTCAACATCCGCTACCTGAGAGAAATGATCGGAGTGGTCAGCCAGGAGCCCGTCTTTTTCCGCCACCACA
TCACCGAGAACATCAGATACGGCCGACTGGACGTGACGCAAGAGGAGATOGAACGAGCCACTAAAGAGTCCAAACGCTTATGA
CTTCATCATGAACCTTCCAGACAAGTTTGAGACGCTGGTGGGAGATCGAGGGACTCAGCTGAGCGGAGGACAGAAGCAGAGG
ATCGCCATOGCTCGAGCTCTGGTCCGCAACCTAAAATCCTCCTGCTGGACGAGGCCACGCTGCOCTOGATGCTGAGAGCG
AGACCATCGTACAGGCTGCTCTGGACAAGGTCGACTGGGTGGCACCACCATCGTCACTCGCTCACCGACTCTOGACCATCAG
AAAACCGACATCATTTGCTGGATTGCTAATGGTGAATCGTTGAGCAGGGGACTCACAGCCAGCTGATGGAGATAAAGGGA
GTCTATCATGGCTGGTGACCATGCAGAGCTTTCAGAAGCTGGAGGATCTGGAAGACTCAGACTAOGAGCCCTGGGTGCTG
AGAAGAGCCAGCTGATCGAATCCTTCTCCAGTCTCCCTGCAGAGGAGGAGTCCACTAGAGGCTCCTTGTCTGCTGCTC
AGAAGGAACAAAAGAGGAGAAAGAAAAATTTGAGTGGATCAGGACAACATAGAGGAGGATGAGAACGTTCTCCCGTGTG
TTCTTTAAAGTGATGGGTTACAACGTTTCTGAGTGGCCGTATATTTGGTAGGAACCATCTGCGCATGATCAACGGTGGGA
TGCAGCCAGTGTTCAGCATCATCTTACCGAGATCATTATGGTCTTCAGAGAAAAGGATAAAGAAATTATCCGGGAGAAGTC
GTCAATCTTTGTATACTGTTTGTGTGATGGGGTGGTGACTTTTCTCACCATGTTTCTGCAGTTTTGGGGTTTCCAGGGT
TTCTGTTTCAGTAAATCTGGAGAAATCTGACCTGAACCTCAGACTCAAAGCCTTCATATCTATGATGAGACAGGACCTCA
GCTGGTACGACAATCCAAAAACCGTTGGCGCTCTCACCACTAGGCTGGCTGCGACGCGCCACGTAACAGGAGCTGC
AGGGGTGGCCTGGCTGTAATGACGAGAACCTGCAACCTGGGCACCAGCATCATCAGCTTCGTGTACGGCTGGGAG
CTGACCTGCTCATCTGGCGTGGTGCCATCTGGCTGTGGCCGGAGCCGCTGAGGTCAAGCTGCTGACAGGACAACGCG
CCGAGACAAGAAGGAGCTGGAGATGGCCGAAAGATCGCCACAGAGGCCATCGAGAATGTGAGAAGTGTGGTGTCCCTCAC
CAGAGAACCGACATTTGTGGCTTTATACGAGGAAAATCTAACTGTTCCATACAAGAAGTCCAGAAAAAGGCCAAAATTTAT
GGCTTAACCTACTCCTTCTCACAGGCCATGATCTTCTTTGTTTACGCTGCTGTTTCCGCTTTGGAGCCTGGCTGATCGAAG

CAGGACGGATGGATGTGGAGGGAGTGTTCCTTGTGGTTATGACAATGCTGTACGGGCGCATGGCTGTGGCGAGGCCAACAC
 TTATGCTCCCAACTTOGCCAAAGCCAAAATCTCAGCCTCCACCTGACGATGCTAATAAACAGACAGCCGCCATAGATAAT
 CTGTCAGAGGAGGAAGCGAGACTGGAGAAATACGACGGCAACGTTCTTTTTGAGGACGTCAAGTTTAACTACCCGTGGCGG
 CCGATGTGCCTGTACTACAAGGGCTGAATCTGGAGGTGCAAAAGGGAGAACTCTGGCCTTGGTGGGCAGCAGCGGTTGTGG
 AAAGAGCACCAACATCCAGCTGCTGGAGAGGTTTTATGACCCAGAGAGGGGAGAGTGTGCTGGACGGTGTGATGTGAAA
 CAGCTGAACGTTCACTGGCTGAGGTCTCAGATCGGCATCGTCTCCAGGAGCCGGTGTGTTGACTGCTCCCTGGCTGAGA
 ACATGCTACGGAGACAACAGTGGCTCCGTGTCATGGATGAGATAGTAGCTGTGCTAAAGCAGCCAACATCCACAGCTT
 CATOGAAGGGCTGCCTCAGAGATACGACACTCAGGCTGGTGATAAGGGAACACAGCTGTCAGGGGGCCAGAAGCAGCGTGT
 GCCATAGCCCGAGCCATCATCCGCAACCCCAAAGTGTGCTCCTGGACGAGGCCACGTCTGGCCTGACACTGAGAGTGAGA
 AGGTGGTGCAGGAGCGTGGACCAGGCCAGGAAGGCGCAGGACGTGCATCGTGTAGCCACCGTCTGTCCACCATCCAGAA
 CGCCGACTGCATCGCTGTGTTCCAGGGAGGAGTGGTGGTGGAAAAGGGGACGCACCAGCAGCTGATCGCCAAGAAGGGAGTG
 TACCACATGCTGGTCACCAACAGATGGGCTATCACAGTGG

1bp-14kb 的预测 cDNA 的氨基酸序列

FRFADGWDILMVTIGVLMIAIVNGLVNPLMCIIVFGEMTDSFIQEAKLSQNHINTSNPTSTMSNVLRVCCIIRHLFSIYYSILGF
 AVLVVAYLQMSLWTLTAARQAKRIRELFHGMQDDISWYDVTETGELNTRLTEWEGIGDKAGLLIQAASTFITSFVIGFVH
 GWKLTILVILAISPVGLLSAALYSKLLTSFTSKEQTAYAKAGAVAAEVLSSIRTVFAFSGQRKAIKRYHKNLEDARDMGIKKG
 VAANTATGFSFLMIYLSYALAFWYGTTLVLNKEYTIGNLLTVFFVLYGAYIGQASPVQSFASARGAAAYKVYNIIDHKPN
 IDSFSEDGYKPEYIKGDIVFQNIHFSYPSRPEIKILNDSMFHVRNGQTIALVGSAGCGKSTTIQLLQRFYDPQKGSIFIDGH
 DIRSLNIRYLREMIGVVSQEPVLFATTITENIRYGRLDVTQEEIERATKESNAYDFIMNLPDKFETLVGDRGTQLSGGQKQR
 IAARALVRNPKILLDEATSALDAESETIVQAALDKVRLGRTTIVIAHRLSTIRNADIAGFSNGEIVEQGTHSQLMEIKG
 VYHGLVTMQSFQKLEDESDYEPVVAEKSQLESFSSQLQRRRSTRGSL LAVSEGTEKEKEKFECDQDNIEEDENPPVS
 FFKVMRYNVSEWPYILVGTICAMINGAMQPVFSIIFTEIIMVFREKDKETIREKSSFFCILFAVMGVVTFMTFLQFWGFG
 FCFKSKGEILTLNLRKAFISMQRDLSWYDNPKNVTGALTRLAADAHVQGAAGVRLAVMTQNFANLGTSIISFVYQWE
 LTLILAVVPILAVAGAAEVKLLTGHAEDKKELEMAGKIATEAENVRTVVSLTREPTFVALYEENLTPYKNSQKAKIY
 GLTYSFSQAMIFFVYAACFRFGAWLIEAGRMDVEGVFLVMTMLYGAMAVGEANTYAPNFAKAKISASHLTMLINRQPAIDN
 LSEEEARLEKYDGNVLFEDVKFNYPSPDPVVLQGLNLEVQKGETLALVGSAGCGKSTTIQLLERFYDPREGRVLLDGVVVK
 QLNVHWLRSQIGIVSQEPVLFDCSLAENIAYGDNRSVSMDEIVAAKAANIHSFIEGLPQRYDTQAGDKGTQLSGGQKQRV
 AIAARAIIRNPKLLLLDEATSALDTESEKVVQALDQARKGRTCIVVAHRLSTIQNADCIAVFQGGVVVEKGTHTQLIAKGV
 YHMLVTKQMGYHSG

12kb-26kb 的预测 cDNA 的序列

TTCCGATTTGCAGACCGCTGGGACGTGCTGCTCATCAGCGGACAGTGTGGCCATGGTCAACGGCAACCGTATGCCCC
 TCATGTGCATTGTCTTTGGAGAAATGACGGACAGTTTTATATACGCTGACATGGCCCAACACAACGCAAGTGGCTGGAATC
 TACTACTACTATTCTGAACAGCAGCTTACAGGAGGACATGCAAGATTCCGCAATTTACTCOGCTCTGGATTGTGTGTG
 CTGCTGGCCGCTACATGCAGGTGTCCTTCTGGACCATAACAGCCGGGCCAGGTGAAACGCATCCGACGCTGTTTTTCC
 ACTGCATCATGCAGCAGGAGATCAGCTGGTTGACGTGAACGACACAGGGGAGCTCAACACTCGACTGACGGACGACGTCTA

CAAGATCCAGGAGGGCATCGGTGACAAGGTGGGGCTGCTGATCCAGGCGTACACCACCTTCATCAOAGGCTTCATCATCGGC
TTCACCAOAGGCTGGAACTGACGCTGGTCACTCTGGCCGTGAGCCCGCGCTGGCCATTTCCGGCGCTTCTTCAGTAAAG
TGCTTGCGTCTTCACCAGTAAGGAGCAGACGGGTACGCCAAAGCCGGAGCGGTGGGGAGGAAGTGTGTCCGCCATCAG
GACCGTGTTCGCTTCAGTGGTCAGACCAGAGAGATTGAGAGATACCACAAGAACCCTGGGGAGCAAAGGACGTGGGAGTG
AAGAAGGCCATCTCTCCAACATCGCCATGGGCTTCACTTCTGTATGATCTACCTGTCTATGCTCTGGCTTCTGGTACG
GGAGTAGCTCATCTGAATTTTGGAGTACCCATCGGCAATTTACTGACTGTGTTTTTTGTGTGCTTATTGGAGCGTTCAG
CGTGGACAGAOCTCTCCGAACATCCAGAATTTTCCAGGCGCCGAGGAGCCCTATAAAGTGTACAGCATCATCGATAAC
AAGCAAACATTGACAGCTTTTCAGAGGACGGTTTCAAGCCGACTTTCATCAAAGGTGACATCGAGTTCAAGAACATCCACT
TCAATTACCTTGGAGGCTGAAGTCAAAAATCTTGAACAACATGTCTCTGAGCGTGAAGAGCGGACAGACCATTGCTTTGGT
GGGAGCAGCGCTGTGGCAAAGTACCACCATCCAGCTGTGCAGAGGTTCTACGACCCCGAGGAAGGAGCTGTATTTATC
GACGGTCAAGACATCCGCTCCCTCAACATCCGCTACCTGAGAGAGATGATCGGAGTGGTCAGCCAGGAGCCGTTCTTTTGG
CCACCAACATCACCGAGAACATCAGATAACGCGGACTGGACGTGACGCAAGAGGAGATCGAACGAGCCACTAAAGAGTCCAA
CGCTTATGACTTCATCATGAACCTTCCAGACAAGTTTGGAGCGTGGTGGGAGATCGAGGGACTCAGCTGAGCGGAGGACAG
AAGCAGAGGATCGCCATCGCTCGAGCTCTGGTCCGCAACCTAAAAATCTCTGCTGGACGAGGCCACGTCTGCCCTCGATG
CTGAGAGCGAGACCATCGTACAGGCTGCTCTGGACAAGTCCGACTGGGTGCAACCAACATCGTGGTCCGCTCACCGACTCTC
GACCATCAGAAAGCCGACATCATTGCTGGATTAGTAATGGCAAATCGTGGAGCAGGGGACTCACAGCCAGCTGATGGAG
ATAAAGGGAGTCTATCATGGCTGGTGACCATGCAGAGTTCACAAATGTGGAGGAGGAAAATACCGCATGTCGGAGTTAT
CTGCTGGGGAGAAGAGCCCTGTGGAAAAGACCGTCTCCAGTCGTCCATCATCAGGAGGAAGTCCACCAGAGGGTCTCTGTT
TGCCGCGTCAGAAGGAACCAAAGAGGAAAAGACAGAAGAGGATGAAGACGTTCCCGAGTGTGTTCTTTAAAGTGTGCAT
CTGAACATCCCGAGTGGCCCTACATCCTTGTGGGGCTCATCTGGCTACGATCAATGGAGCCATGCAGCCGCTCTTCCCA
TCTCTTCTCCAAGATCATCACTGTGTTTGGGATCCAGACCGTATTCTGTGTCAGGAGGAAGAGTGAATTCATTTCTCTGAT
GTTTGTGTTATTGGCTGTGTGCAATTTGTACCATGTTTTTACAGGGTACTGTTTCCGTAATCCCGAGAGATTCTGACG
CTGAAGCTGAGACTCCGGGGTTACGGCGATGATGAGACAGGACCTCAGCTGGTAGACAATCCCAAAAACCGTTGGCG
CTCTCAACACTAGGCTGGCTGCCGACGCCGCCAAGTACAAGGAGCTGCAGGGGTGGCGCTGGGACAATAATGCAGAACTT
CGCAAACCTGGGACACCAGCATCATATCGCTTTGTTAAGGCTGGGAGCTGACCTGCTCATCTGGCGTGGTGGCCCTC
ATCGCGCCCGCGAGCCGCTGAGATCAAGCTGCTCGCGGTCAAGCCGCAAGACAAGAAGGAGCTGGAGAAGGCCGAA
AGATCGCCACAGAGGCCATCGAGAACGTCAGAACCGTGTGTCCTCAGCAGAGAACCAAAATTTGAGTGTATATGAGGA
GAATCTCAGAGTCCGTTACAAGAACTCCAGAAAAAGGCCACGTTACGGCTTAACCTACTCTCTCCAGGCCATGATC
TACTTTGCTTACGCTGCTGTTTCCGCTTCCGAGCCTGGCTGATTGAAGCAGGACGGATGGAAGTGGAGGGAGTGTCTCTGG
TGGTTTCTGGGTGCTGTACGGCGCCATGGCCGTGGGGGAAGCTAACACCTTTGCTCCGAACCTACGCCAAGGCCAAAAATGGC
TGCTTCTACCTGATGATGCTAATAAACAAGAAGCCCGCATTTGATAACCTCTCAGAGGAGGGGACGTTCTCCGAAAAATAC
GACGGTAATGTGCATTTCCAGGGTGTAAATTTCAACTACCGTCCGCGCCGATGTGACCATACTCCAGGGGCTGAACCTGA
AGGTGAAAAAGGGAGAACTCTGGCCTTGGTGGGCAGCAGCGTTGTGGAAAGAGCAACCATCCAGCTGCTGGAGAGGTT
TTATGACCCAGAGAGGGGAGAGTGTCACTGGACGGTGTCAACGTGAAACAGCTGAACATTCAGTGGCTGAGGCTCAGATC
GGCATCGTCTCCAGGAGCCGCTGCTGTTCCGACTGCTCCCTGGCTGAGAACATCGCTACGGAGACAACAGTCCGCTCCGCTG

CCATGGATGAGATAGTACTTGCTGCTAAAGCAGCCAACATCCACAGCTTCATCGAAGGGCTGCCTCAGAGATACGACACTCA
GGCTGGTGATAAGGGAACACAGCTGTGAGGGGGCCAGAAGCAGCGTGTGCCATAGCCCGAGCCATCATCCGCAACCCCAAA
CTGTTGCTCCTGGACGAGGCCAAGTCTGCGCTCGACACTGAGAGTGAGAAGGTGGTGCAGGAGGCGTTGGACCAGGCCAGGA
AGGGCAGGACGTGCATCGTGGTAGCCACCGTCTGTCCACCATCCAGAACGCCGACTGCATCGCTGTGTTCCAGGGAGGAGT
GGTGGTGAAAAGGGGACGCACCAGCAGCTGATCGCCAAGAAGGGAGGTACCACATGCTGGTCAACAAACAGATGGGCTAT
CACAACGAC

12kb-26kb 的预测 cDNA 的氨基酸序列

FRFADRWDVLLISGTVMAMVNGTVMPLMCIVFGEMTDSFIYADMAQHNASGWNSTTTILNSTLQEDMQRFAYIYYSVLGFVV
LLAAYMQVSWFTITAGRQVKRIRSLFFHCIMQQEISWFDVNDTGLNTRLTDDVYKIQEGIGDKVGLLIQAYTTFITAFIIG
FTTGWKLTLVILAVSPALAISSAFFSKVLASFTSKEQTAYAKAGAVAEVLSAIRTVFAFSGQREIERYHKNLRDAKDVGV
KKAISSNIAMGFTFLMIYLSYALAFWYGSTLILNFEYTIENLLTVFFVVLIGAFSVGQTSPIQNFASARGAAYKVYSIIDN
KPNIDFSSEDGFKPDFIKGDIKFNIFNYPSPRPEVKILNMSLSVKSGQTIALVSSGCGKSTTIQLLQRFYDPEEGAVFI
DGHDIRSLNIRYLREMIQVVSQEPVLFATTITENIRYGRLDVTQEEIERATKESNAYDFIMNLPDKFETLVGDRGTQLSGGQ
KQRIATARALVRNPKILLDEATSALDAESETIVQAALDKVRLGRTTIVVAHRLSTIRNADI IAGFSNGKIVEGQTHSQLME
IKGVYHGLVTMQTFHNVEEENTAMSELSAGEKSPVEKTVSQSSI IRRKSTRGSSFAASEGKKEEKEEDEDVDPVDFKVLH
LNIPWPYILVGLICATINGAMQPVFAIFLSKIITVFADPDRDSVRRKSEFISLMFVVICVSVFVTFMQGYCFGKSGEILT
LKLRLRAFTAMMRQDLWYDNPQNTVYGALTTTLAADAQVQGAAGVRLATIMQNFANLGTSSII IAFVYGWELTLLILAVVPL
IAAGAAEIKLLAGHAADKKELEKAGKIATEA IENVRTVVSLSREPKEFLYEENLRVPYKNSQKKAHVYGLTYSFSQAMI
YFAYAACFRFGAWLIEAGRMDVEGVFLVVSAVLYGAMAVGEANTFAPNYAKAKMAASYLMLINKKPAIDNLSEEGTSPEKY
DGNVHFEGVKFNYPSPRPDVTILQGLNLKVKKGETLALVSSGCGKSTTIQLLERFYDPREGRVSLDGVNVKQLNIHWLRSQI
GIVSQEPVLFDCSLAENIAYGDNRSRVSMDIIVLAAKAANIHSFIEGLPQRYDTQAGDKGTQLSGGQKQVARIARAIIRNPK
LLLLDEATSALDTESEKVVQALDQARKGRTCI VVAHRLSTIQNADCI AVFQGGVVVEKGT HQQLI AKKGVYHMLVTKQMGY
HND

利用 blastx 软件对 fuguMDR1 和 fuguMDR2 进行同源性比较, 结果证实:

fuguMDR1 cDNA 与人的 MDR1 cDNA 的氨基酸同源性为 75%

Sequences producing significant alignments:	Score	E
	(bits)	Value
sp P21448 MDR1_CRIGR MULTIDRUG RESISTANCE PROTEIN 1 (P-GLYCOPRO...	1463	0.0
sp P08183 MDR1_HUMAN MULTIDRUG RESISTANCE PROTEIN 1 (P-GLYCOPRO...	1460	0.0
sp P21439 MDR3_HUMAN MULTIDRUG RESISTANCE PROTEIN 3 (P-GLYCOPRO...	1445	0.0
sp P21447 MDR3_MOUSE MULTIDRUG RESISTANCE PROTEIN 3 (P-GLYCOPRO...	1430	0.0

sp P06795 MDR1_MOUSE	MULTIDRUG RESISTANCE PROTEIN 1 (P-GLYCOPRO...	1425	0.0
sp P23174 MDR3_CRIGR	MULTIDRUG RESISTANCE PROTEIN 3 (P-GLYCOPRO...	1425	0.0
sp P21440 MDR2_MOUSE	MULTIDRUG RESISTANCE PROTEIN 2 (P-GLYCOPRO...	1419	0.0
sp P21449 MDR2_CRIGR	MULTIDRUG RESISTANCE PROTEIN 2 (P-GLYCOPRO...	1417	0.0
sp Q08201 MDR2_RAT	MULTIDRUG RESISTANCE PROTEIN 2 (P-GLYCOPROTE...	1411	0.0
sp P43245 MDR1_RAT	MULTIDRUG RESISTANCE PROTEIN 1 (P-GLYCOPROTE...	1404	0.0
sp P34712 MDR1_CAEEL	MULTIDRUG RESISTANCE PROTEIN 1 (P-GLYCOPRO...	940	0.0
sp Q00449 MDR4_DROME	MULTIDRUG RESISTANCE PROTEIN HOMOLOG 50 (P...	899	0.0
sp Q00748 MDR5_DROME	MULTIDRUG RESISTANCE PROTEIN HOMOLOG 65 (P...	883	0.0
sp P36619 PMD1_SCHPO	LEPTOMYCIN B RESISTANCE PROTEIN PMD1	824	0.0
sp P34713 MDR3_CAEEL	MULTIDRUG RESISTANCE PROTEIN 3 (P-GLYCOPRO...	815	0.0
sp Q06034 MDR1_LEIEN	MULTIDRUG RESISTANCE PROTEIN 1 (P-GLYCOPRO...	689	0.0
sp P13568 MDR_PLAFF	MULTIDRUG RESISTANCE PROTEIN (CHLOROQUINE R...	491	e-138
sp P12866 STE6_YEAST	MATING FACTOR A SECRETION PROTEIN STE6 (MU...	417	e-116
sp P53706 HST6_CANAL	ATP-DEPENDENT PERMEASE HST6 (STE6 HOMOLOG)	303	1e-81
sp P44407 MSBA_HAEIN	PROBABLE TRANSPORT ATP-BINDING PROTEIN MSBA	289	3e-77
sp P27299 MSBA_ECOLI	PROBABLE TRANSPORT ATP-BINDING PROTEIN MSBA	284	9e-76
sp P33310 MDL1_YEAST	ATP-DEPENDENT PERMEASE MDL1	278	5e-74
sp P97998 MDL1_CANAL	ATP-DEPENDENT PERMEASE MDL1	277	9e-74
sp P33311 MDL2_YEAST	ATP-DEPENDENT PERMEASE MDL2	263	2e-69
sp P22638 HETA_ANASP	HETEROCYST DIFFERENTIATION ATP-BINDING PRO...	259	2e-68
sp P36372 TAP2_RAT	ANTIGEN PEPTIDE TRANSPORTER 2 (APT2)	256	1e-67
sp P45861 YWJA_BACSU	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING P...	256	2e-67
sp P11599 HLYB_PROVU	HEMOLYSIN SECRETION ATP-BINDING PROTEIN	251	5e-66
sp P26760 RT1B_ACTPL	RTX-I TOXIN DETERMINANT B (TOXIN RTX-I SEC...	250	8e-66
sp P16532 HLYB_PASHA	LEUKOTOXIN SECRETION ATP-BINDING PROTEIN	250	1e-65
sp P36371 TAP2_MOUSE	ANTIGEN PEPTIDE TRANSPORTER 2 (APT2) (HIST...	249	3e-65
sp P36370 TAP1_RAT	ANTIGEN PEPTIDE TRANSPORTER 1 (APT1)	248	4e-65
sp P21958 TAP1_MOUSE	ANTIGEN PEPTIDE TRANSPORTER 1 (APT1) (HIST...	248	4e-65
sp P55122 HLYB_PASSP	LEUKOTOXIN SECRETION ATP-BINDING PROTEIN	247	9e-65
sp P54719 YFIC_BACSU	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING P...	246	2e-64
sp P10089 HLY2_ECOLI	HEMOLYSIN SECRETION ATP-BINDING PROTEIN, C...	246	2e-64
sp Q04473 RT3B_ACTPL	RTX-III TOXIN DETERMINANT B (TOXIN RTX-III...	245	4e-64
sp P08716 HLYB_ECOLI	HEMOLYSIN SECRETION ATP-BINDING PROTEIN, P...	245	5e-64
sp P55469 Y4GM_RHISN	PROBABLE ABC TRANSPORTER ATP-BINDING PROTE...	243	2e-63

sp P23702 HLYB_ACTAC	LEUKOTOXIN SECRETION ATP-BINDING PROTEIN	243	2e-63
sp Q03518 TAP1_HUMAN	ANTIGEN PEPTIDE TRANSPORTER 1 (APT1) (PEPT...	240	2e-62
sp P40416 ATM1_YEAST	MITOCHONDRIAL TRANSPORTER ATM1 PRECURSOR	230	1e-59
sp P47261 Y015_MYCGE	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING P...	229	2e-59
sp Q02592 HMT1_SCHPO	HEAVY METAL TOLERANCE PROTEIN PRECURSOR	228	5e-59
sp Q03519 TAP2_HUMAN	ANTIGEN PEPTIDE TRANSPORTER 2 (APT2) (PEPT...	224	6e-58
sp P75094 Y015_MYCPN	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING P...	224	1e-57
sp Q11047 Y08A_MYCTU	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING P...	224	1e-57
sp Q57180 YA51_HAEIN	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING P...	218	6e-56
sp P18768 CHVA_AGRTU	BETA-(1->2)GLUCAN EXPORT ATP-BINDING PROT...	217	1e-55
sp P18767 NDVA_RHIME	BETA-(1->2)GLUCAN EXPORT ATP-BINDING PROT...	213	2e-54
sp Q23868 TAGC_DICDI	PRESTALK-SPECIFIC PROTEIN TAGC PRECURSOR	212	4e-54
sp P54683 TAGB_DICDI	PRESTALK-SPECIFIC PROTEIN TAGB PRECURSOR	210	1e-53
sp P18770 CYAB_BORPE	CYCLOLYSIN SECRETION ATP-BINDING PROTEIN CYAB	210	1e-53
sp P54718 YFIB_BACSU	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING P...	201	8e-51
sp Q11018 Y07C_MYCTU	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING P...	184	9e-46
sp P30751 MDL_ECOLI	MULTIDRUG RESISTANCE-LIKE ATP-BINDING PROTE...	182	3e-45
sp Q03727 COMA_STRPN	TRANSPORT ATP-BINDING PROTEIN COMA	180	1e-44
sp P36497 PEDD_PEDAC	PEDIOCIN PA-1 TRANSPORT ATP-BINDING PROTEI...	178	7e-44
sp Q11046 Y089_MYCTU	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING P...	174	8e-43
sp Q11019 Y07D_MYCTU	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING P...	173	2e-42
sp Q57538 Y664_HAEIN	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING P...	171	9e-42
sp Q00564 LCNC_LACLA	LACTOCOCCIN A TRANSPORT ATP-BINDING PROTEI...	166	3e-40
sp Q10418 MESD_LEUME	MESENTERICIN Y105 TRANSPORT ATP-BINDING PR...	164	7e-40
sp P22520 CVAB_ECOLI	COLICIN V SECRETION ATP-BINDING PROTEIN CVAB	161	1e-38
sp P94367 CYDD_BACSU	TRANSPORT ATP-BINDING PROTEIN CYDD	156	3e-37
sp P23886 CYDC_ECOLI	TRANSPORT ATP-BINDING PROTEIN CYDC	154	7e-37
sp P29018 CYDD_ECOLI	TRANSPORT ATP-BINDING PROTEIN CYDD	146	2e-34
sp P33116 SPAT_BACSU	SUBTILIN TRANSPORT ATP-BINDING PROTEIN SPAT	146	2e-34
sp P75095 Y014_MYCPN	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING P...	146	2e-34
sp P33527 MRP1_HUMAN	MULTIDRUG RESISTANCE-ASSOCIATED PROTEIN 1	146	2e-34
sp P23596 PRTD_ERWCH	PROTEASES SECRETION ATP-BINDING PROTEIN PRTD	146	3e-34
sp P47260 Y014_MYCGE	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING P...	146	3e-34
sp P45081 CYDC_HAEIN	TRANSPORT ATP-BINDING PROTEIN CYDC	144	1e-33
sp P45082 CYDD_HAEIN	TRANSPORT ATP-BINDING PROTEIN CYDD	139	2e-32

sp Q10185 YAWB_SCHPO	PROBABLE ATP-DEPENDENT PERMEASE C3F10.11C	139	2e-32
sp Q03024 APRD_PSEAE	ALKALINE PROTEASE SECRETION ATP-BINDING PR...	138	6e-32
sp P38735 YHD5_YEAST	PROBABLE ATP-DEPENDENT PERMEASE YHL035C	138	8e-32
sp P14772 YL15_YEAST	PROBABLE ATP-DEPENDENT TRANSPORTER YLL015W	136	2e-31
sp Q09429 SUR_RAT	SULFONYLUREA RECEPTOR	136	3e-31
sp P94366 CYDC_BACSU	TRANSPORT ATP-BINDING PROTEIN CYDC	136	3e-31
sp Q09428 SUR_HUMAN	SULFONYLUREA RECEPTOR	134	7e-31
sp Q09427 SUR_CRICR	SULFONYLUREA RECEPTOR	134	1e-30
sp P16875 MDR1_ENTHI	MULTIDRUG RESISTANCE PROTEIN 1 (P-GLYCOPRO...	133	2e-30
sp P39109 YCF1_YEAST	METAL RESISTANCE PROTEIN YCF1 (YEAST CADMI...	132	5e-30
sp P53049 YOR1_YEAST	OLIGOMYCIN RESISTANCE ATP-DEPENDENT PERMEA...	131	8e-30
sp Q92337 ABC1_SCHPO	ATP-BINDING CASSETTE TRANSPORTER ABC1	130	1e-29
sp P16877 MDR4_ENTHI	MULTIDRUG RESISTANCE PROTEIN 4 (P-GLYCOPRO...	130	2e-29
sp P30750 ABC_ECOLI	ATP-BINDING PROTEIN ABC	127	2e-28
sp P16876 MDR3_ENTHI	MULTIDRUG RESISTANCE PROTEIN 3 (P-GLYCOPRO...	126	3e-28
sp P47650 PSTB_MYCGE	PHOSPHATE TRANSPORT ATP-BINDING PROTEIN PS...	122	3e-27
sp P75186 PSTB_MYCPN	PHOSPHATE TRANSPORT ATP-BINDING PROTEIN PS...	122	3e-27
sp Q50294 Y179_MYCPN	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING P...	118	5e-26
sp P37608 LCN3_LACLA	LACTICIN 481/LACTOCOCCIN TRANSPORT ATP-BIN...	117	1e-25
sp P45171 POTA_HAEIN	SPERMIDINE/PUTRESCINE TRANSPORT ATP-BINDIN...	116	2e-25
sp P44785 ABC_HAEIN	ATP-BINDING PROTEIN ABC	116	3e-25
sp Q03203 NIST_LACLA	NISIN TRANSPORT ATP-BINDING PROTEIN NIST	116	3e-25
sp Q58418 PSTB_METJA	PROBABLE PHOSPHATE TRANSPORT ATP-BINDING P...	115	4e-25
sp P47425 Y179_MYCGE	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING P...	114	1e-24
sp P31134 POTG_ECOLI	PUTRESCINE TRANSPORT ATP-BINDING PROTEIN POTG	113	2e-24
sp P54537 YQIZ_BACSU	PROBABLE AMINO-ACID ABC TRANSPORTER ATP-BI...	111	1e-23

sp|P08183|MDR1_HUMAN MULTIDRUG RESISTANCE PROTEIN 1 (P-GLYCOPROTEIN 1)

Length = 1280

Score = 1460 bits (3739), Expect = 0.0

Identities = 735/1240 (59%), Positives = 940/1240 (75%), Gaps = 14/1240 (1%)

Query: 1 FRFADGWDILMVTIGVLMIAIVNGLVNLPLMCIVFGEMTDSFIQEAKLSQ-----NHNTSN 162
FR+++ D L + +G L AI++G PLM +VFGEMTD F L N + N

Sbjct: 40 FRYSNWLDKLYMVVGTAAI IHGAGLPLMMLVFGEMTDIFANAGNLEDLMSNITNRSIN 99
 Query: 163 PTSTMSNVLRVCCIIRHLFSIYYSILGFAVLVAYLQMSLWTLTAARQAKRIRELFPHGI 342
 T N+ ++ YYS +G VLV AY+Q+S W L A RQ +IR+ FFH I

Sbjct: 100 DTGFFMNLEEDMT—RYAYYSGIGAGVLVAAYIQVSFWCLAAGRQIHKIRKQFFHAI 155
 Query: 343 MQQDISWYDVTETGELNTRLTEW—EGIGDKAGLLIQAASTFITSFVIGFVHGWKLT 507
 M+Q+I W+DV + GELNTRLT+ EGIGDK G+ Q+ +TF T F++GF GWKLT

Sbjct: 156 MRQEIGWFDVHDVDELNTRLTDDVSKINEGIGDKIGMFFQSMATFFTFIVGFTRGWKLT 215
 Query: 508 LVILAI SPVLGLSAA LYSKLLTSFTSKEQTXXXXXXXXXXEVLSSIRTVFAFSGQRKAIK 687
 LVILAI SPVLGLSAA+++K+L+SFT KE EVL++IRTV AF GQ+K ++

Sbjct: 216 LVILAI SPVLGLSAAVWAKILSSFTDKELLAYAKAGAVAEVLA AIRTVIAFGGQKKELE 275
 Query: 688 RYHKNLEDARDMGIKKGVAANTATGFSFLMIYLSYALAFWYGTTLVLNKEYTIGNLLTVF 867
 RY+KNLE+A+ +GIKK + AN + G +FL+IY SYALAFWYGTTLV+ EY+IG +LTVF

Sbjct: 276 RYNKNLEEAKRIGIKKAITANISIGAAFLIYASYALAFWYGTTLVLSGEYSIGQVLTVF 335
 Query: 868 FVVLYGAYIIGQASPNVQSFASARGAAYKVYNIIDHKPNIDSFSEDGYKPEYIKGDIVFQ 1047
 F VL GA+ +QASP++++FA+ARGAAY+++ IID+KP+IDS+S+ G+KP+ IKG++ F+

Sbjct: 336 PSVLIGAFSVGQASPSIEAFANARGAAYEIFKIIDNKPSIDSYSKSGHKPDNIKGNLEFR 395
 Query: 1048 NIHFYSYPSRPEIKILNDSFHRNGQTIALVGSSGCGKSTTIQLLQRFYDPOKGSIFIDG 1227
 N+HFYSYPSR E+KIL ++ V++GQT+ALVG+SGCGKSTT+QL+QR YDP +G + +DG

Sbjct: 396 NVHFYSYPSRKEVKILKGLNLKVGSGQTVALVGNVSGCGKSTTVQLMQRLYDPTEGMVSDG 455
 Query: 1228 HDIRSLNIRYLREMIGVVSQEPVLFATTITENIRYGRLDVTQEEIERATKESNAYDFIMN 1407
 DIR++N+R+LRE+IGVVSQEPVLFATTI ENIRYGR +VT +EIE+A KE+NAYDFIM

Sbjct: 456 QDIRTINVRFLREIIGVVSQEPVLFATTIAENIRYGRENVTMDEIEKAVKEANAYDFIMK 515
 Query: 1408 LPDKFETLVGDRGTQLSGGQKQRIAIARALVRNPKILLLDEATSALDAESETIVQAALDK 1587
 LP KF+TLVG+RG QLSGGQKQRIAIARALVRNPKILLLDEATSALD ESE +VQ ALDK

Sbjct: 516 LPHKFDTLVGERGAQLSGGQKQRIAIARALVRNPKILLLDEATSALDTESEAVVQVALDK 575
 Query: 1588 VRLGRTTIVIAHRLSTIRNADI IAGFSNGEIVEQGTHSQLMEIKGVYHGLVTMQSFQKLE 1767
 R GRTTIVIAHRLST+RNAD+IAGF +G IVE+G H +LM+ KG+Y LVTMQ+

Sbjct: 576 ARKGRTTIVIAHRLSTVRNADVIAGFDDGVIVEKGNHDELMKEKGIYFKLVMTQTAGNEV 635
 Query: 1768 DLEDSDYEPWVAEKSQL--IEXXXXXXXXXXXXXXXXXXAVSEGTKEEKEKFECDQDNIE 1941
 +LE++ E KS++ +E +V G++ + K +

Sbjct: 636 ELENADE—SKSEIDALEMSSNDSRSSLIRKRSTRRSV-RGSQAQDRKLTKE—A 687
 Query: 1942 EDENVPPVSFFKVMRYNVSEWPYILVGTICAMINGAMQPVFSIIFTEIIMVF-REKDKEI 2118
 DE++PPVSF+++M+ N++EWPY +VG CA+ING +QP F+IIF++II VF R D E

Sbjct: 688 LDESIPPVSFWRIMKLNLTWPYFVVGVFCAIINGGLQAFAIIFSKIIGVFTRIDDPET 747
 Query: 2119 IREKSSFFCILFAVMGVVFTLTMFLQFWGFGFCFSKSGEILTLNRLKAFISMRRQDLS 2298

R+ S+ F +LF +G+++F+T FLQ GF F K+GEILT LR F SM+RQD+S
 Sbjct: 748 KRQNSNLFSLFLALGIISFITFFLQ——GFTFGKAGEILTKRLRYMVFRSMLRQDVS 802
 Query: 2299 WYDNPKNVTGALTTRLXXXXXXXXXXXXRLAVMTQNFANLGTSIISFVYGWEXXXXXX 2478
 W+D+PKNT GALTTRL RLAV+TQN ANLGT IISF+YGW+
 Sbjct: 803 WFDDPKNTTGTALTRLANDAAQVKAIGSRLAVITQNIANLGTGIIISFIYGWQLTLLLL 862
 Query: 2479 XXXXXXXXXXXXEVKLLTGHAEDKKELEMAGKIATEAIENVRTVVSILTREPTFVALYEE 2658
 E+K+L+G A +DKKELE AGKIATEAIEN RTVVSLT+E F +Y +
 Sbjct: 863 AIVPIIAIAGVEMKMLSGQALKDKKELEGAGKIATEAIENFRTVVSLTQEQKFEHMYAQ 922
 Query: 2659 NLTPYKNSQKKAKIYGLTYSFSQAMIFFVYAACFRFGAWLIEAGRMDVEGVFLVVMTML 2838
 +L VPY+NS +KA I+G+T+SF+QAM++F YA CFRFGA+L+ M E V LV ++
 Sbjct: 923 SLQVPYRNSLRKAHIFGITPSFTQAMMYFSYAGCFRFGAYLVAHKLMSFEDVLLVFSAVV 982
 Query: 2839 YGAMAVGEANTYAPNFAKAKISASHLTM LINRQPAIDNLSEEEARLEKYDGNVLFEDVKF 3018
 +GAMAVG+ +++AP++AKAKISA+H+ M+I + P ID+ S E +GNV F +V F
 Sbjct: 983 FGAMAVGQVSSFAPDYAKAKISAAHIIMIIEKTPLIDSYSTEGLMPNTLEGNVTFGEVVF 1042
 Query: 3019 NYPSRPDPVVLQGLNLEVKGETLALVGSSGCGKSTTIQLLERFYDPREGRVLLDGVVDK 3198
 NYP+RPD+PVLQGL+LEV+KG+TLALVGSSGCGKST +QLLERFYDP G+VLLDG ++K
 Sbjct: 1043 NYPTRPDIPVLQGLSLEVKKGQTLALVGSSGCGKSTVQLLERFYDPLAGKVLDDGKEIK 1102
 Query: 3199 QLNVHWLRSQIGIVSQEPVLFDCSLAENIAYGDNSRSVSMDEIVAAAKAANIHSFIEGLP 3378
 +LNV WLR+ +GIVSQEP+LFDCS+AENIAYGDNSR VS +EIV AAK ANIH+FIE LP
 Sbjct: 1103 RLVNQWLRHLGIVSQEPILFDCSLAENIAYGDNSRVVSQEEIVRAAKEANIHFIESLP 1162
 Query: 3379 QRYDTQAGDKGTQLSGGQKQRAIARAIIRNPKLLLLDEATSALDTESEKVVQEALDQAR 3558
 +Y T+ GDKGTQLSGGQKQRAIARA++R P +LLLDEATSALDTESEKVVQEALD+AR
 Sbjct: 1163 NKYSTKVGDKGTQLSGGQKQRIARALVRQPHILLLDEATSALDTESEKVVQEALDKAR 1222
 Query: 3559 KGRTCIVVAHRLSTIQNADCIAVFQGGVVVEKGTHTQQLIAKKGVYHMLVTKQMG 3720
 +GRTCIV+AHRLSTIQNAD I VFQ G V E GTHQQL+A+KG+Y +V+ Q G
 Sbjct: 1223 EGRTCIVIAHRLSTIQNADLIVVFQNGRVKEHGTHQQLAQLKGIYFSMVSQAG 1276

fuguMDR2 cDNA 与人的 MDR1 cDNA 的氨基酸同源性的为 76%

Sequences producing significant alignments:	Score	E
	(bits)	Value
sp P08183 MDR1_HUMAN MULTIDRUG RESISTANCE PROTEIN 1 (P-GLYCOPRO...	1524	0.0
sp P21448 MDR1_CRIGR MULTIDRUG RESISTANCE PROTEIN 1 (P-GLYCOPRO...	1512	0.0
sp P06795 MDR1_MOUSE MULTIDRUG RESISTANCE PROTEIN 1 (P-GLYCOPRO...	1501	0.0

sp P21447 MDR3_MOUSE	MULTIDRUG RESISTANCE PROTEIN 3 (P-GLYCOPRO...	1498	0.0
sp P21439 MDR3_HUMAN	MULTIDRUG RESISTANCE PROTEIN 3 (P-GLYCOPRO...	1490	0.0
sp P43245 MDR1_RAT	MULTIDRUG RESISTANCE PROTEIN 1 (P-GLYCOPROTE...	1481	0.0
sp P21449 MDR2_CRIGR	MULTIDRUG RESISTANCE PROTEIN 2 (P-GLYCOPRO...	1475	0.0
sp P21440 MDR2_MOUSE	MULTIDRUG RESISTANCE PROTEIN 2 (P-GLYCOPRO...	1465	0.0
sp P23174 MDR3_CRIGR	MULTIDRUG RESISTANCE PROTEIN 3 (P-GLYCOPRO...	1462	0.0
sp Q08201 MDR2_RAT	MULTIDRUG RESISTANCE PROTEIN 2 (P-GLYCOPROTE...	1459	0.0
sp P34712 MDR1_CAEEL	MULTIDRUG RESISTANCE PROTEIN 1 (P-GLYCOPRO...	961	0.0
sp Q00449 MDR4_DROME	MULTIDRUG RESISTANCE PROTEIN HOMOLOG 50 (P...	912	0.0
sp Q00748 MDR5_DROME	MULTIDRUG RESISTANCE PROTEIN HOMOLOG 65 (P...	904	0.0
sp P34713 MDR3_CAEEL	MULTIDRUG RESISTANCE PROTEIN 3 (P-GLYCOPRO...	831	0.0
sp Q06034 MDR1_LEIEN	MULTIDRUG RESISTANCE PROTEIN 1 (P-GLYCOPRO...	719	0.0
sp P13568 MDR_PLAFF	MULTIDRUG RESISTANCE PROTEIN (CHLOROQUINE R...	484	e-136
sp P36619 PMD1_SCHPO	LEPTOMYCIN B RESISTANCE PROTEIN PMD1	469	e-131
sp P12866 STE6_YEAST	MATING FACTOR A SECRETION PROTEIN STE6 (MU...	434	e-121
sp P30751 MDL_ECOLI	MULTIDRUG RESISTANCE-LIKE ATP-BINDING PROTE...	316	2e-85
sp P53706 HST6_CANAL	ATP-DEPENDENT PERMEASE HST6 (STE6 HOMOLOG)	308	4e-83
sp P44407 MSBA_HAEIN	PROBABLE TRANSPORT ATP-BINDING PROTEIN MSBA	302	3e-81
sp P27299 MSBA_ECOLI	PROBABLE TRANSPORT ATP-BINDING PROTEIN MSBA	298	3e-80
sp P33311 MDL2_YEAST	ATP-DEPENDENT PERMEASE MDL2	295	3e-79
sp P33310 MDL1_YEAST	ATP-DEPENDENT PERMEASE MDL1	293	1e-78
sp P97998 MDL1_CANAL	ATP-DEPENDENT PERMEASE MDL1	287	1e-76
sp P22638 HETA_ANASP	HETEROCYST DIFFERENTIATION ATP-BINDING PRO...	277	1e-73
sp P45861 YWJA_BACSU	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING P...	267	1e-70
sp P36372 TAP2_RAT	ANTIGEN PEPTIDE TRANSPORTER 2 (APT2)	266	2e-70
sp P36370 TAP1_RAT	ANTIGEN PEPTIDE TRANSPORTER 1 (APT1)	265	4e-70
sp P21958 TAP1_MOUSE	ANTIGEN PEPTIDE TRANSPORTER 1 (APT1) (HIST...	263	1e-69
sp P36371 TAP2_MOUSE	ANTIGEN PEPTIDE TRANSPORTER 2 (APT2) (HIST...	261	7e-69
sp P55469 Y4GM_RHISN	PROBABLE ABC TRANSPORTER ATP-BINDING PROTE...	258	4e-68
sp Q03518 TAP1_HUMAN	ANTIGEN PEPTIDE TRANSPORTER 1 (APT1) (PEPT...	256	2e-67
sp P54719 YFIC_BACSU	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING P...	255	3e-67
sp P10089 HLY2_ECOLI	HEMOLYSIN SECRETION ATP-BINDING PROTEIN, C...	254	9e-67
sp P08716 HLYB_ECOLI	HEMOLYSIN SECRETION ATP-BINDING PROTEIN, P...	254	9e-67
sp P55122 HLYB_PASSP	LEUKOTOXIN SECRETION ATP-BINDING PROTEIN	252	4e-66
sp P11599 HLYB_PROVU	HEMOLYSIN SECRETION ATP-BINDING PROTEIN	252	4e-66

sp P16532 HLYB_PASHA	LEUKOTOXIN SECRETION ATP-BINDING PROTEIN	251	5e-66
sp P26760 RT1B_ACTPL	RTX-I TOXIN DETERMINANT B (TOXIN RTX-I SEC...	250	1e-65
sp P23702 HLYB_ACTAC	LEUKOTOXIN SECRETION ATP-BINDING PROTEIN	248	5e-65
sp Q04473 RT3B_ACTPL	RTX-III TOXIN DETERMINANT B (TOXIN RTX-III...	246	2e-64
sp Q11047 Y08A_MYCTU	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING P...	238	6e-62
sp P47261 Y015_MYCGE	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING P...	238	6e-62
sp Q03519 TAP2_HUMAN	ANTIGEN PEPTIDE TRANSPORTER 2 (APT2) (PEPT...	236	2e-61
sp P75094 Y015_MYCPN	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING P...	235	3e-61
sp P54718 YFIB_BACSU	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING P...	228	4e-59
sp Q02592 HMT1_SCHPO	HEAVY METAL TOLERANCE PROTEIN PRECURSOR	227	1e-58
sp P40416 ATM1_YEAST	MITOCHONDRIAL TRANSPORTER ATM1 PRECURSOR	225	4e-58
sp P18767 NDVA_RHIME	BETA-(1->2)GLUCAN EXPORT ATP-BINDING PROT...	215	5e-55
sp P18768 CHVA_AGRU	BETA-(1->2)GLUCAN EXPORT ATP-BINDING PROT...	212	2e-54
sp Q57180 YA51_HAEIN	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING P...	211	6e-54
sp Q23868 TAGC_DICDI	PRESTALK-SPECIFIC PROTEIN TAGC PRECURSOR	210	2e-53
sp P54683 TAGB_DICDI	PRESTALK-SPECIFIC PROTEIN TAGB PRECURSOR	208	6e-53
sp P18770 CYAB_BORPE	CYCLOLYSIN SECRETION ATP-BINDING PROTEIN CYAB	207	1e-52
sp Q11018 Y07C_MYCTU	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING P...	186	2e-46
sp Q11046 Y089_MYCTU	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING P...	186	2e-46
sp P36497 PEDD_PEDAC	PEDIOCIN PA-1 TRANSPORT ATP-BINDING PROTEI...	185	3e-46
sp Q11019 Y07D_MYCTU	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING P...	182	5e-45
sp Q03727 COMA_STRPN	TRANSPORT ATP-BINDING PROTEIN COMA	180	1e-44
sp Q57538 Y664_HAEIN	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING P...	176	2e-43
sp P22520 CVAB_ECOLI	COLICIN V SECRETION ATP-BINDING PROTEIN CVAB	171	5e-42
sp Q00564 LCNC_LACLA	LACTOCOCCIN A TRANSPORT ATP-BINDING PROTEI...	169	3e-41
sp Q10418 MESD_LEUME	MESENTERICIN Y105 TRANSPORT ATP-BINDING PR...	169	3e-41
sp P47260 Y014_MYCGE	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING P...	166	3e-40
sp P75095 Y014_MYCPN	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING P...	163	1e-39
sp P94367 CYDD_BACSU	TRANSPORT ATP-BINDING PROTEIN CYDD	155	4e-37
sp P23886 CYDC_ECOLI	TRANSPORT ATP-BINDING PROTEIN CYDC	154	7e-37
sp P29018 CYDD_ECOLI	TRANSPORT ATP-BINDING PROTEIN CYDD	150	1e-35
sp P33527 MRP1_HUMAN	MULTIDRUG RESISTANCE-ASSOCIATED PROTEIN 1	148	6e-35
sp P39109 YCF1_YEAST	METAL RESISTANCE PROTEIN YCF1 (YEAST CADMI...	146	3e-34
sp P45082 CYDD_HAEIN	TRANSPORT ATP-BINDING PROTEIN CYDD	146	3e-34
sp P33116 SPAT_BACSU	SUBTILIN TRANSPORT ATP-BINDING PROTEIN SPAT	144	1e-33

sp P38735 YHD5_YEAST	PROBABLE ATP-DEPENDENT PERMEASE YHL035C	141	7e-33
sp Q10185 YAWB_SCHPO	PROBABLE ATP-DEPENDENT PERMEASE C3F10.11C	141	7e-33
sp P45081 CYDC_HAEIN	TRANSPORT ATP-BINDING PROTEIN CYDC	141	7e-33
sp P23596 PRTD_ERWCH	PROTEASES SECRETION ATP-BINDING PROTEIN PRTD	141	7e-33
sp P94366 CYDC_BACSU	TRANSPORT ATP-BINDING PROTEIN CYDC	138	5e-32
sp Q03024 APRD_PSEAE	ALKALINE PROTEASE SECRETION ATP-BINDING PR...	138	6e-32
sp P16875 MDR1_ENTHI	MULTIDRUG RESISTANCE PROTEIN 1 (P-GLYCOPRO...	132	5e-30
sp P14772 YL15_YEAST	PROBABLE ATP-DEPENDENT TRANSPORTER YLL015W	131	1e-29
sp P16877 MDR4_ENTHI	MULTIDRUG RESISTANCE PROTEIN 4 (P-GLYCOPRO...	130	2e-29
sp P75186 PSTB_MYCPN	PHOSPHATE TRANSPORT ATP-BINDING PROTEIN PS...	129	2e-29
sp Q92337 ABC1_SCHPO	ATP-BINDING CASSETTE TRANSPORTER ABC1	128	7e-29
sp Q09429 SUR_RAT	SULFONYLUREA RECEPTOR	128	7e-29
sp Q09428 SUR_HUMAN	SULFONYLUREA RECEPTOR	128	7e-29
sp Q09427 SUR_CRICR	SULFONYLUREA RECEPTOR	127	1e-28
sp P30750 ABC_ECOLI	ATP-BINDING PROTEIN ABC	127	2e-28
sp P53049 YOR1_YEAST	OLIGOMYCIN RESISTANCE ATP-DEPENDENT PERMEA...	125	4e-28
sp P16876 MDR3_ENTHI	MULTIDRUG RESISTANCE PROTEIN 3 (P-GLYCOPRO...	124	8e-28
sp P47650 PSTB_MYCGE	PHOSPHATE TRANSPORT ATP-BINDING PROTEIN PS...	122	5e-27
sp P44785 ABC_HAEIN	ATP-BINDING PROTEIN ABC	119	3e-26
sp P37608 LCN3_LACLA	LACTICIN 481/LACTOCOCCIN TRANSPORT ATP-BIN...	118	6e-26
sp P45171 POTA_HAEIN	SPERMIDINE/PUTRESCINE TRANSPORT ATP-BINDIN...	118	8e-26
sp Q03203 NIST_LACLA	NISIN TRANSPORT ATP-BINDING PROTEIN NIST	115	7e-25
sp P54537 YQIZ_BACSU	PROBABLE AMINO-ACID ABC TRANSPORTER ATP-BI...	114	9e-25
sp Q50294 Y179_MYCPN	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING P...	114	1e-24
sp Q58418 PSTB_METJA	PROBABLE PHOSPHATE TRANSPORT ATP-BINDING P...	113	2e-24
sp Q45460 PROV_BACSU	GLYCINE BETAINE/L-PROLINE TRANSPORT ATP-BI...	111	1e-23
sp P47425 Y179_MYCGE	HYPOTHETICAL ABC TRANSPORTER ATP-BINDING P...	110	1e-23

sp|P08183|MDR1_HUMAN MULTIDRUG RESISTANCE PROTEIN 1 (P-GLYCOPROTEIN 1)

Length = 1280

Score = 1524 bits (3903), Expect = 0.0

Identities = 759/1229 (61%), Positives = 951/1229 (76%), Gaps = 10/1229 (0%)

Query: 1 FRFADRWDVLLISGTVMAMVNGTVMPLMCIVFGEMTDSFIYADMAQH-----NASGWN 162
FR+++ D + ++ GT+ A+++G +PLM +VFGEMTD F A + N S N

Sbjct: 40 FRYSNWLDKLYMVVGTAAIIHGAGLPLMMLVFGEMTDIFANAGNLEDLMSNITNRSIN 99
Query: 163 STTTILNSTLQEDMORFAIYYSVLGFVLLAAYMQVSWTITAGRQVKRIRSLFFHCIMQ 342
T +N L+EDM R+A YYS +G VL+AA+QVSW + AGRQ+ +IR FFH IM+

Sbjct: 100 DTGFFMN—LEEDMTRYAYYSYGIGAGVLAAYIQVSWCLAAGRQIHKIRKQFFHAIMR 157
Query: 343 QEISWFDVNDTGELNRLTDDVYKIQEIGDKVGLLIQAYXXXXXXXXXXXXXXXXXWKLTLV 522
QEI WFDV+D GELNRLTDDV KI EGIGDK+G+ Q+ WKLTLV

Sbjct: 158 QEIGWFDVHDVDELNRLTDDVSKINEGIGDKIGMFFQSMATFFTFGIVGFTRGWKLTLV 217
Query: 523 ILAVSPALISAFFSKVLASFTSKEQTAYAKAGAVAEVLSAIRTVFAPSGQREIERY 702
ILA+SP L +SAA ++K+L+SFT KE AYAKAGAVAEVL+AIRTV AF GQ +E+ERY

Sbjct: 218 ILAISPVLGLSAAVWAKILSSFTDKELLAYAKAGAVAEVLAIRTVIAFGGQKKELEERY 277
Query: 703 HKNLDAKDVGVKKAISSNIAMGFTFLMIYLSYALAFWYGSTLILNFEYTIENLLTVFFV 882
+KNL +AK +G+KKA++NI++G FL+IY SYALAFWYG+TL+L+ EY+IG +LTVFF

Sbjct: 278 NKNLEEAKRIGIKKAITANISIGAAFLLIYASYALAFWYGTTLVLSGEYSIGQVLTVFFS 337
Query: 883 VLIGAFSVGQTSPNIQNFASARGAAYKVYSIIDNKPIDSFSSEDFKPDFIKGDIEFKNI 1062
VLIGAFSVGQ SP+I+ FA+ARGAAY+++ IIDNKP+IDS+S+ G KPD IKG++EF+N+

Sbjct: 338 VLIGAFSVGQASPSIEAFANARGAAYEIFKIIDNKPIDSYSKSGHKPDNIKGNLEFRNV 397
Query: 1063 HFNYPSRPEVKILNMSLSVKSGQTIALVGSSGCGKSTTIQLLQRFYDPEEGAVFDGHD 1242
HF+YPSR EVKIL ++L V+SGQT+ALVG+SGCGKSTT+QL+QR YDP EG V +DG D

Sbjct: 398 HFSYPSRKEVKILKGLNLKVQSGQTVALVNSGCGKSTTVQLMQLYDPTGEMVSDGQD 457
Query: 1243 IRSLNIRYLREMIGVVSQEPVLFATTITENIRYGRDVTQEEIERATKESNAYDFIMNLP 1422
IR++N+R+LRE+IGVVSQEPVLFATTI ENIRYGR +VT +EIE+A KE+NAYDFIM LP

Sbjct: 458 IRTINVRFLREIIGVVSQEPVLFATTIAENIRYGRENVMTDEIEKAVKEANAYDFIMKLP 517
Query: 1423 DKFETLVGDRGTQLSGGQKORIAIARALVRNPKILLLDEATSALDAESETIVQAALDKVR 1602
KF+TLVG+RG QLSGGQKORIAIARALVRNPKILLLDEATSALD ESE +VQ ALDK R

Sbjct: 518 HKFDTLVGERGAQLSGGQKORIAIARALVRNPKILLLDEATSALDTESEAVVQVALDKAR 577
Query: 1603 LGRTTIVVAHRLSTIRNADIIAGFSNGKIVEGTHSQLMEIKGVYHGLVTMQTFHN-VEE 1779
GRTTIV+AHRLST+RNAD+IAGF +G IVE+G H +LM+ KG+Y LVTMQT N VE

Sbjct: 578 KGRTTIVIAHRLSTVRNADVIAGFDDGVIVEKGNHDELMKEKGIYFKLVMTQTAGNEVEL 637
Query: 1780 ENTAMSELSAGEKSPVEKTVSQSSIIRKSTRGS—SFAASEGXXXXXXXXXXVPDVSF 1953
EN A S + + S+SS+IR++STR S A + +P VSF

Sbjct: 638 ENAADESKSEIDALEMSSNDSRSSLIRKRSTRRSVRGSAQDRKLSKEALDESIPVVSF 697
Query: 1954 FKVHLNIPWPYILVGLICATINGAMQPVFAILFSKIITVFAD-PDRDSVRRKSEFISL 2130
+++ LN+ EWPY +VG+ CA ING +QP FAI+FSKII VF D ++ R+ S SL

Sbjct: 698 WRIMKLNLTWPYFVVGVFCAIINGLQPAFAIIFSKIIGVFTRIDDPETKRQNSNLFSL 757
Query: 2131 MFVVIGCVSFVTFMLQCYCFGKSGEILTLKRLRAFTAMMRQDLSWYDNPQNTVGALTTX 2310

+F+ +G +SF+T FLOG+ FGK+GEILT +LR F +M+RQD+SW+D+P+NT GALT
 Sbjct: 758 LFLALGIISFITFFLOGFTFGKAGEILTKRLRYMVFMSMLRQDVSWFDDPKNTTGALTTR 817
 Query: 2311 XXXXXXXXXXXXXXXXXXXTIMQNFANLGTSLIIAFVYGWEXXXXXXXXXXXXXXXXXXXXX 2490
 I QN ANLGT III+F+YGW+

Sbjct: 818 LANDAAQVKAIGSRLAVITQNIANLGTGIIISFIYQWQLTLLLLAIVPIIAIAGVEMK 877
 Query: 2491 XXXGHAAKDKKELEKAGKIATEAIEIVRIVVSLSRPKFECLYEENLRVPYKNSQKKAHV 2670
 G A KDKKELE AGKIATEAIEI RTVVSLE++E KFE +Y ++L+VPY+NS +KAH+

Sbjct: 878 MLSCQALKDKKELEGAGKIATEAIEIVRIVVSLTQEQKFEHMYAQSLOVYPYRNSLRKAHI 937
 Query: 2671 YGLTYSFSQAMIYFAYAACFRFGAWLIEAGRMDVEGVFLVVSALYVAVGAMAVGEANTFAPN 2850
 +G+T+SF+QAM+YF+YA CFRFGA+L+ M E V LV SAV++GAMAVG+ ++FAP+

Sbjct: 938 FGITFSFTQAMMYFSYAGCFRFGAYLVAHKLSFEDVLLVFSAVVFGAMAVGQVSSFPAD 997
 Query: 2851 YAKAKMAASYLMLINKPAIDNLSEEGTSPEKYDGNVHFEGVKFNYPSPRPDVTILQGLN 3030
 YAKAK++A+++M+I K P ID+ S EG P +GNV F V FNYP+RPD+ +LQGL+

Sbjct: 998 YAKAKISAAHIIMIEKTPLIDSYSTEGLMPNTLEGNTFGEVVFNYTPRPDIPVLQGLS 1057
 Query: 3031 LKVKKGETLALVSSGCGKSTTIQLLERFYDPREGRVSLDGVNVKQLNIHWLRSQIGIVS 3210
 L+VKKG+TLALVSSGCGKST +QLLERFYDP G+V LDG +K+LN+ WLR+ +GIVS

Sbjct: 1058 LEVKKGQTLALVSSGCGKSTVQQLLERFYDPLAGKVLLDGKEIKRLNVQWLRHLGIVS 1117
 Query: 3211 QEPVLFDCSLAENIAYGDNSRSVSMDEIVLAAKAANIHSFIEGLPQRYDTQAGDKGTQLS 3390
 QEP+LFDCS+AENIAYGDNSR VS +EIV AAK ANIH+FIE LP +Y T+ GDKGTQLS

Sbjct: 1118 QEPILFDCSIAENIAYGDNSRVVSQEEIVRAAKEANIHFIESLPNKYSTKVGDKGTQLS 1177
 Query: 3391 GGQQRVAIARAIIRNPKLLLLDEATSALDTESEKVVQEALDQARKGRTCIVVAHRLSTI 3570
 GGQQR+AIARA++R P +LLLLDEATSALDTESEKVVQEALD+AR+GRTCIV+AHRLSTI

Sbjct: 1178 GGQQRVAIARALVRQPHILLDEATSALDTESEKVVQEALDKAREGRTCIVIAHRLSTI 1237
 Query: 3571 QNADCIAVFQGGVVVEKGTHTQQLIAKKGVYHMLVTKQMG 3687
 QNAD I VFQ G V E GTHQQL+A+KG+Y +V+ Q G

Sbjct: 1238 QNADLIVVFQNGRVKEHGTHQQLLAQKGIYFSMVSQAG 1276

III 河豚鱼 *fuguMDR1* cDNA 和 *fuguMDR2* cDNA 的核苷酸结构、氨基酸一级、二级结构分析

使用 Dnastar 的 Megalign 程序对 *fuguMDR1* cDNA 和 *fuguMDR2* cDNA 与多种生物的多药耐药基因或 ATP-结合蛋白进行比较, 结果如下:

多种生物的多药耐药基因之间核苷酸水平的同源性比较

	MDLE-D	

	10 20 30 40	
mdr1FUGU.seq	-----	1
mdr2FUGU.seq	-----	1
mdr1CRIG.seq	MEFEEDFSG-RK--DK-----	40
mdr1HUMA.seq	MDLEGDRNGGAK--KK-----	43
mdr3MOUS.seq	MELEEDLKG-RA--DK-----	40
mdr2-RAT.seq	MDLEAARNGTARRLDG-----	49
mdr1CAEE.seq	MLRNGSLRQSLRTLDSFSL-----	58
mdr4DROM.seq	MVKKEESR--LPQAGDFQL-----	52
pmd1SCHP.seq	MSLHSHKSTSTVKDNEHSLDLSIKSIPSNKNEKNEFSTKSEN	118
mdr1LEIE.seq	M-----SRAHAAYANEGWSA	43
mdrPLAFF.seq	MGKEQKEKKDG-----NLSIK-----	49
ste6YEAS.seq	MNF-----LSFKTTKHYH-----	40
msbaECOL.seq	MHNDKDLST-----WQT	34
hlybPROV.seq	MDFHKK-----	19
	-----K-----	
	50 60 70 80	
mdr1FUGU.seq	-----	1
mdr2FUGU.seq	-----	1
mdr1CRIG.seq	-----NFLKMGKRS-----K--KEKKEKK-	91
mdr1HUMA.seq	-----NFFKLNNKS-----E--KDKKEKK-	94
mdr3MOUS.seq	-----NFSKMGKKS-----K--KEKKEKK-	91
mdr2-RAT.seq	-----DF-ELGSIS-----NQSREKKKKV-	103
mdr1CAEE.seq	--APEDVLKTAIKTVEDYEGDNIDSNGEIKITRDAKEEVV	169
mdr4DROM.seq	--KEGSVVDATRK---Y-----	88
pmd1SCHP.seq	EASESHVVDVVKDPFEQYTPPEEQEI-----LYKQINDTPA	223
mdr1LEIE.seq	PEGGIAGKDGSTRDCSGYGSQGPLFSAE---EEVKGTVV	151
mdrPLAFF.seq	-----EEVEKELNKKKSTAELEFRKIK	106
ste6YEAS.seq	-----	40
msbaECOL.seq	-----	37
hlybPROV.seq	-----	19
	-----S-----FRYADWLDRLLLLVGTLAAIIHGAGLPLMML	
	90 100 110 120	
mdr1FUGU.seq	-----FRFADGWDILMVTIGVLMIAIVNGLVNPLMCI	91
mdr2FUGU.seq	-----FRFADRDVLLISGTVMMAMVNGTVMPLMCI	91
mdr1CRIG.seq	P-VVSVFTMFRYAGWLDRLYMLVGTLAAIIHGVALPLMML	205
mdr1HUMA.seq	P-TVSVFMSFRYSNWLDKLYMVVGTLAAIIHGAGLPLMML	208
mdr3MOUS.seq	P-AVSVLTMFRYAGWLDRLYMLVGTLAAIIHGVALPLMML	205
mdr2-RAT.seq	N-LIGPLTLFRYSDWQDKLFMLLGTAMAI AHGSGLPLMMI	217
mdr1CAEE.seq	N-KVSI PQLYRYTTTLEKLLLFIGTLVAVITGAGLPLMSI	286
mdr4DROM.seq	----SYFDLFRYSTRCERFLLVVSLLVATAASAFAIPYFMI	193
pmd1SCHP.seq	K-LSGYPRI LSYADKWDIMLQLAGTITGIGAGLGMPPLMSL	340
mdr1LEIE.seq	RETVGPIEIFRYADATDRVLMIAGTAFAVACGAGMPVFSF	271
mdrPLAFF.seq	NEKISFFLPPFKCLPAQHRKLLFISFVCAVLSGGTLPPFFIS	226
ste6YEAS.seq	-----IFRYVNI RNDYRL-----LMIM	88
msbaECOL.seq	-----FR-----RLWPTIAPFKAGLIVAGVAL	100
hlybPROV.seq	-----	19

VFGEMTDSFA---GNIGT--N-TL-----N-T-----
 130 140 150 160

mdr1FUGU.seq VFGEMTDSFI-----QEAKLSQNH-----NTSNPTST-MS 178
 mdr2FUGU.seq VFGEMTDSFI-----YADMAQHNA-----SGWNSTTTILN 181
 mdr1CRIG.seq VFGDMTDSFA-SVGNIPT--NATN-----NATQVNASDIF 301
 mdr1HUMA.seq VFGEMTDIFA-NAGNLEDLMSNIT-----NRSDINDTGFF 310
 mdr3MOUS.seq IFGDMTDSFA-SVGNVSK--NSTN-----MSEADKRAMF 298
 mdr2-RAT.seq VFGEMTDKFDVDNAGNFSLPVPNFSL-----SMLNPGRI-- 316
 mdr1CAEE.seq LQGKVSQAFINEQIVINNNGSTFL-----PTGQNYTKTDF 391
 mdr4DROM.seq IYGEFTSLLVDRTVGVGTSSPAFALPMFGGGQQLTNASKE 313
 pmd1SCHP.seq VSGQLAQAFDLDASGKGASS-----FQHTVD----- 421
 mdr1LEIE.seq IFGRIAMDLMGVSAGSAEKAAKTSLIM----- 355
 mdrPLAFF.seq VFGVILKNM-----NLGDDINPIILSL----- 295
 ste6YEAS.seq IIGTVATGLVPAITSILTGRVFDLLSVFVA--NGSHQGLY 202
 msbaECOL.seq ILNAASDTFMLSLLK-----PLLDDGFGKTDORSVL 190
 hlybPROV.seq ----- 19

--L---E-MTRYAYYSYGIGAGVL--IAAYIQVSFWTLA
 170 180 190 200

mdr1FUGU.seq NVLRVCCIIRHLFSIYYSILGFAVL--VVAYLQMSLWTLT 292
 mdr2FUGU.seq STLQ---EDMQRFAIYYSVLGFVVL--LAAYMQVSFWTIT 286
 mdr1CRIG.seq GKLE---EEMTTYAYYYTGIGAGVL--IVAYIQVSFWCLA 406
 mdr1HUMA.seq MNLE---EDMTRYAYYSYGIGAGVL--VAAYIQVSFWCLA 415
 mdr3MOUS.seq AKLE---EEMTTYAYYYTGIGAGVL--IVAYIQVSFWCLA 403
 mdr2-RAT.seq --LE---EEMTRYAYYSGLGGGVL--LAAYIQVSFWTLA 412
 mdr1CAEE.seq E-----H.DVMNVVWSYAAMTVGMW--AAGQIIVTCYLYV 487
 mdr4DROM.seq ENNQAIIDDATAFGIGSLVGSVAMF--LLITLAIDLANRI 427
 pmd1SCHP.seq -----HFCLYFIYIAIGVF--GCSYIYTVTFIIA 499
 mdr1LEIE.seq -----VYVGIAML--IACAGHVMCWTVA 415
 mdrPLAFF.seq -----VSIGLVQFILSMISSYCMDVI 355
 ste6YEAS.seq SQLVQRSMAVMALG---AASVPVMW---LSLTSWMHI 295
 msbaECOL.seq VWMP-----VIGLMILRGITSYVSSYCIISWV 274
 hlybPROV.seq -----NNYGLYALEI-----LAQYHNISI--- 76

AGRQIKRIRQLFFHAILRQE-IGWFDVNDT--GELNTRLT
 210 220 230 240

mdr1FUGU.seq AARQAKRIRELFFHGMQQD-ISWYDVTET--GELNTRLT 403
 mdr2FUGU.seq AGRQVKRIRSLFFHCIMQQE-ISWFDVNDT--GELNTRLT 397
 mdr1CRIG.seq AGRQIHKIRQKFFHAIMNQE-IGWFDVHDV--GELNTRLT 517
 mdr1HUMA.seq AGRQIHKIRKQFFHAIMRQE-IGWFDVHDV--GELNTRLT 526
 mdr3MOUS.seq AGRQIHKIRQKFFHAIMNQE-IGWFDVHDV--GELNTRLT 514
 mdr2-RAT.seq AGRQIRKIRQKFFHAILRQE-MGWFDIKGT--TELNTRLT 523
 mdr1CAEE.seq AEQMNNRLRREFVKSILRQE-ISWFDTNHS--GTLATKLF 598
 mdr4DROM.seq ALNQIDRIRKLFLEAMLRQD-IAWYDTSSG--SNFASKMT 538
 pmd1SCHP.seq GERIARRIRQDYLAHAILSQN-IGYFDRLGA--GEITTRIT 610
 mdr1LEIE.seq ACRQVARIRLLFFRAVLRQD-IGWHDEHSP--GALTARMT 526
 mdrPLAFF.seq TSKILKTLKLEYLRSVIFYQD--GQFHDNPNP--SKLRSDDL 466
 ste6YEAS.seq GERQGFRIRSQILEAYLEEKPMEWYDNNEKLLGDF-TQIN 412
 msbaECOL.seq SGKVVMTMRRRLFGHMMGMP-VSFFDKQST--GTLLSRIT 385
 hlybPROV.seq ---NPEEIKHKF-----D-I----- 109

DDVSKIN EGIGDKIGMFFQAIATFIGGFIIGFTRGWKLT L
 250 260 270 280

mdr1FUGU.seq E - - - - WEGIGDKAGLLIQAASTFITSFVIGFVHGWKLT L 508
 mdr2FUGU.seq DDVYKIQEGIGDKVGLLIQAYTTFITAFIIGFTTGWKLT L 517
 mdr1CRIG.seq DDVSKIN EGIGDKIGMFFQAMATFFGGFIIGFTRGWKLT L 637
 mdr1HUMA.seq DDVSKINEVIGDKIGMFFQSMATFFTGFIVGFTRGWKLT L 646
 mdr3MOUS.seq DDVSKIN EGIGDKIGMFFQAMATFFGGFIIGFTRGWKLT L 634
 mdr2-RAT.seq DDISKISEGIGDKVGMFFQAIATFFAGFIVGFIRGWKLT L 643
 mdr1CAEE.seq DNLERVKEGTGDKIGMAFYLSQFITGFIVAFTHSWQLT L 718
 mdr4DROM.seq EDLDK LKEGIGEKIVIVVFLIMTFVIGIVSAFVYGWKLT L 658
 pmd1SCHP.seq TDTNFIQDGLGEKVG L VFFAIATFVSGFVIAFIRHWKFT L 730
 mdr1LEIE.seq GDTRVIQNGINDKLSQGITMNGSMGVIGYIAGFVFSWELT L 646
 mdrPLAFF.seq FYLEQVSSGIGTKFITITFYASSFLGLYIWSLIK NARLT L 586
 ste6YEAS.seq RCVEELRSSSAEASAITFQNLVAICALLGTSFYYSWSLT L 532
 msbaECOL.seq YDSEQVASSSSGALITV VREGASIIIGL FIMMFYYSWQLS I 505
 hlybPROV.seq - - - - - NGVGLDLT - - - - - SWLLAA 148

VILAISPVLGLSAAVWAKILSSFTDKELLAYAKAGAVAAE
 290 300 310 320

mdr1FUGU.seq VILAISPVLGLSAAALYSKLLTSFTSKEQTAYAKAGAVAAE 628
 mdr2FUGU.seq VILAVSPALAISAAFFSKVLASFTSKEQTAYAKAGAVAAE 637
 mdr1CRIG.seq VILAISPVLGLSAGIWAKILSSFTDKELQAYAKAGAVAAE 757
 mdr1HUMA.seq VILAISPVLGLSAAVWAKILSSFTDKELLAYAKAGAVAAE 766
 mdr3MOUS.seq VILAISPVLGLSAGIWAKILSSFTDKELHAYAKAGAVAAE 754
 mdr2-RAT.seq VIMAITAILGLSTAVWAKILSTFSDKELAAAYAKAGAVAAE 763
 mdr1CAEE.seq VMLAVTPIQALCGFAIAKSMSTFAIRETLRYAKAGKVVEE 838
 mdr4DROM.seq VVLSVFPFIIAATS VVARLQGS LAEKELKSYS DAANVVEE 778
 pmd1SCHP.seq I LSSMFP AICGGIGLGV PFITKNTKGQI AVVAESSTFVEE 850
 mdr1LEIE.seq MMIGMPPFIIVMAAII GSI VSKITES SRKYFAKAGSLATE 766
 mdrPLAFF.seq CITCVFPLIYVCGVICNKKVKLNKKTSLLYNNNTMSIIEE 706
 ste6YEAS.seq IILCSP IITFFAVVFSRMIHVYSEKENSETSKAAQLLTW 652
 msbaECOL.seq ILIVLAPIV SIAIRV VSKRFRNISKNMQNTMGQVTTSAEQ 625
 hlybPROV.seq K SLEL - - - - - 166

VLSAIRTVIAFGGQKKELE RYNKNLEEA KRLG I KKA I AAN
 330 340 350 360

mdr1FUGU.seq VLSSIRTVFVAFSGQRKAIKRYHKNLE DARDMG I KKGVAAN 748
 mdr2FUGU.seq VLSAIRTVFVAFSGQ TREIER YHKNLRDAKDVGVKKA I SSN 757
 mdr1CRIG.seq VLAAIRTVIAFGGQKKELE RYN N N LEEAKRLG I KKA I TAN 877
 mdr1HUMA.seq VLAAIRTVIAFGGQKKELE RYNKNLEEA KRIG I KKA I TAN 886
 mdr3MOUS.seq VLAAIRTVIAFGGQKKELE RYN N N LEEAKRLG I KKA I TAN 874
 mdr2-RAT.seq ALGAIRTVIAFGGQNKELERYQKHLEN AKKIG I KKA I SAN 883
 mdr1CAEE.seq TISSIRTVVS LNGLRYELERYSTAVEEA KKAGV LKGLFLG 958
 mdr4DROM.seq VFSGIRTVFVAFSGQEKEKERFGKLLI PAENTGRKKGLYSG 898
 pmd1SCHP.seq VFSNIRNAFAFGTQDILAKLYNKYLIT AQRF GINKAIAMG 970
 mdr1LEIE.seq VMENIRTVQAFGRE DYELERFTKAVLYAQGRGIRKELASN 886
 mdrPLAFF.seq ALMGIRTVAS YCGEKTILNKFNLS ETFYSKYILKANFVEA 826
 ste6YEAS.seq SMNAAQLVRLYCTQR LERKKFKEIILNCNTFFIKSCFFVA 772
 msbaECOL.seq MLKGHKEVLI FGGQEVETKRFDK - - - - - 697
 hlybPROV.seq - - - - - KVKAVKKTIER - 199

ISMGA AFLLIYASYALAFWYGTTLVLSK--E-----YTI
 370 380 390 400

mdr1FUGU.seq TATGFSFLMIYLSYALAFWYGTTLVLSK--E-----YTI 844
 mdr2FUGU.seq IAMGFTFLMIYLSYALAFWYGSTLILNF--E-----YTI 853
 mdr1CRIG.seq ISMGA AFLLIYASYALAFWYGTSLVISK--E-----YSI 973
 mdr1HUMA.seq ISIGAAFLLIYASYALAFWYGTTLVLVSG--E-----YSI 982
 mdr3MOUS.seq ISMGA AFLLIYASYALAFWYGTSLVISK--E-----YSI 970
 mdr2-RAT.seq ISMGI AFLLIYASYALAFWYGTSLVISK--E-----YTI 979
 mdr1CAEE.seq ISFGAMQASNFI SFALAFYIGVGVVHDG--S-----LNF 1054
 mdr4DROM.seq MGNALSWLIIYLCMALAIWYGVTLILDE--RDLPDRVYTP 1012
 pmd1SCHP.seq LMVGMFFVAVYGVYGLAFWEGGRL-LHA--GDL-----DV 1066
 mdr1LEIE.seq LSAAVIMALMYVSYTVAFFFGSYLV-----EWGRR--DM 982
 mdrPLAFF.seq LHIGLINGLILVSYAFGFWYGTTRIIINSATNQYPNDFNG 946
 ste6YEAS.seq ANAGILRFLTLTMFVQGFVFGSAMIKKG--K-----LNI 868
 msbaECOL.seq ----- 697
 hlybPROV.seq -----LNFIIYLPALVWRE--DG---HFFIL 256

GDVLT VFFSVLIGAFSLGQASPNIEAFANARGAAYEIFDI
 410 420 430 440

mdr1FUGU.seq GNLLTVFFVVLIGAYIIGQASPNVQSFASARGAAYKVYNI 964
 mdr2FUGU.seq GNLLTVFFVVLIGAFSVGQTSNIIQNIFASARGAAYKVYSI 973
 mdr1CRIG.seq GQVLT VFFFAVLIGAFSIGQASPNIEAFANARGAAYEIFNI 1093
 mdr1HUMA.seq GQVLT VFFSVLIGAFSVGQASPSIEAFANARGAAYEIFKI 1102
 mdr3MOUS.seq GQVLT VFFSVLIGAFSVGQASPNIEAFANARGAAYEVFKI 1090
 mdr2-RAT.seq GNAMTVFFSILIGAFSVGQAAPCIDAFPNARGAAYVIFDI 1099
 mdr1CAEE.seq GDMLTTFSSVMMGSMALGLAGPQLAVLGTAAQGAASGIYEV 1174
 mdr4DROM.seq AVLVI VLFVAVIMGAQNLFASPHVEAIAVATAAGQTLFNI 1132
 pmd1SCHP.seq SKLIGCFFAVLIASYSLANISPKMQSFVSCASAAKKIFDT 1186
 mdr1LEIE.seq ADIISTFLAVLMSYFGLGFVAPSRRTAFTE SRAAYEIFKA 1102
 mdrPLAFF.seq ASVIS ILLGVLI SMFMLTII LPNITEYMKALEATNSLYEI 1066
 ste6YEAS.seq NDVITCFHSCIMLGSTLNNTLHQIVVLQKGGVAMEKIMTL 988
 msbaECOL.seq ----- 697
 hlybPROV.seq TKV-----NKESNRYLIYDL 301

IDRKPSIDSFSKGGHKP---DNIKGDIEFKNVHFSYPSRP
 450 460 470 480

mdr1FUGU.seq IDHKPNIDSFSEGDGYKP---EYIKGDIVFQNIHFSYPSRP 1075
 mdr2FUGU.seq IDNKPNIDSFSEGDGFKP---DFIKGDIEFKNIHFNYPSRP 1084
 mdr1CRIG.seq IDNKPSIDSFSKNGYKP---DNIKGNLEFKNIHFSYPSRK 1204
 mdr1HUMA.seq IDNKPSIDSYSKSGHKP---DNIKGNLEFRNVHFSYPSRK 1213
 mdr3MOUS.seq IDNKPSIDSFSKSGHKP---DNIQGNLEFKNIHFSYPSRK 1201
 mdr2-RAT.seq IDNNPKIDSFSERGHKP---DSIKGNLEFSDVHFSYPSRA 1210
 mdr1CAEE.seq LDRKPVIDSSSKAGRKD---MKIKGDITVENVHFTYPSRP 1285
 mdr4DROM.seq IDRPSQVDPMDEKGNRP---ENTAGHIRFEGIRFRYPARP 1243
 pmd1SCHP.seq IDRVPINAFTPTGDVV---KDIKGEIELKNIRFVYPTRP 1297
 mdr1LEIE.seq IDRVPVD-IDAGGVPV---PGFKESIEFRNVRFAYPTRP 1210
 mdrPLAFF.seq INRKPLVENDDGE---TLPNIK-KIEFKNVRFHVDTRK 1171
 ste6YEAS.seq LKDGSKRNPLNKTVAHQFPLDYATSDLTFANVSFSYPSRP 1108
 msbaECOL.seq ----- 697
 hlybPROV.seq EQRNPRV-----LEQAEFEDL---YQG-- 361

EVKILKGLSLKVKSGQTVALVGSSGCGKSTTIQLLQRFYD

490 500 510 520

mdr1FUGU.seq	EIKILNDMSFHVRNGQTIALVGSSGCGKSTTIQLLQRFYD	1195
mdr2FUGU.seq	EVKILNNMSLSVKSGQTIALVGSSGCGKSTTIQLLQRFYD	1204
mdr1CRIG.seq	DVQILKGLNLKVQSGQTVALVGNSGCGKSTTVQLLQRLYD	1324
mdr1HUMA.seq	EVKILKGLNLKVQSGQTVALVGNSGCGKSTTVQLMQRLYD	1333
mdr3MOUS.seq	EVQILKGLNLKVKSGQTVALVGNSGCGKSTTVQLMQRLYD	1321
mdr2-RAT.seq	NIKILKGLNLKVKSGQTVALVGNSGCGKSTTVQLLQRLYD	1330
mdr1CAEE.seq	DVPILRGMNLRVNAGQTVALVGSSGCGKSTIISLLRYD	1405
mdr4DROM.seq	DVEILKGLTVDVLPGQTVAFVVGASGCGKSTLIQLMQRFYD	1363
pmd1SCHP.seq	EVLVLDNFSLVCPGKITALVGASGSGKSTIIGLVERFYD	1417
mdr1LEIE.seq	GMILFRDLSLKIKCGQKVAFVVGASGCGKSSVIGLIQRFYD	1330
mdrPLAFF.seq	DVEIYKDLSTLKEGKTYAFVVGESGCGKSTILKLIERLYD	1291
ste6YEAS.seq	SEAVLKNVSLNFSAGQFTFIVGKSGSGKSTLSNLLRFYD	1228
msbaECOL.seq	-----	697
hlybPROV.seq	-----NIIILITS-----RSSVIGKLAKFD-	418

PTEGSVSI-DGQDIRTINVRYLREIIGVVSQEPVLFATTI

530 540 550 560

mdr1FUGU.seq	PQKGSIFI-DGHDIRSLNIRYLREMIGVVSQEPVLFATTI	1312
mdr2FUGU.seq	PEEGAVFI-DGHDIRSLNIRYLREMIGVVSQEPVLFATTI	1321
mdr1CRIG.seq	PTEGVVSI-DGQDIRTINVRYLREIIGVVSQEPVLFATTI	1441
mdr1HUMA.seq	PTEGMVSV-DGQDIRTINVRFLREIIGVVSQEPVLFATTI	1450
mdr3MOUS.seq	PLDGMVSI-DGQDIRTINVRYLREIIGVVSQEPVLFATTI	1438
mdr2-RAT.seq	PTEGTISI-DGQDIRNFNVRCLREFIGVVSQEPVLFSTTI	1447
mdr1CAEE.seq	VLK GKITI-DGVDVRDINLEFLRKNVAVVSQEPALFNCTI	1522
mdr4DROM.seq	PEAGSVKL-DGRDLRTLNVGWLRISQIGVVGQEPVLFATTI	1480
pmd1SCHP.seq	PIGGQVFL-DGKDLRTLNVASLRNQISLVQQEPVLFATTV	1534
mdr1LEIE.seq	PIGGAVLV-DGVRMRELCLREWRDQIGIVSQEPNLFAGTM	1447
mdrPLAFF.seq	PTEGDIIVNDSHNLDKINLKWWRISKIGVVSQDPLLFNSNI	1411
ste6YEAS.seq	GYNGSISI-NGHNIQTIDQKLLIENITVVEQRCTLFNDTL	1345
msbaECOL.seq	-----	697
hlybPROV.seq	-----FTWFI	430

AENIRYG-----REDVTMDEIEK-----

570 580 590 600

mdr1FUGU.seq	TENIRYG-----RLDVTQEEIER-----	1369
mdr2FUGU.seq	TENIRYG-----RLDVTQEEIER-----	1378
mdr1CRIG.seq	AENIRYG-----RENVTMDEIEK-----	1498
mdr1HUMA.seq	AENIRYG-----RENVTMDEIEK-----	1507
mdr3MOUS.seq	AENIRYG-----REDVTMDEIEK-----	1495
mdr2-RAT.seq	AENIRYG-----RGNVTMDEIKK-----	1504
mdr1CAEE.seq	EENISLG-----KEGITREEMVA-----	1579
mdr4DROM.seq	GENIRYG-----RPSATQADIEK-----	1537
pmd1SCHP.seq	FENITYGL-----PDTIKGTLKSKEELERRVY-----	1615
mdr1LEIE.seq	MENVRMGK-----PNATDEEVV-----	1501
mdrPLAFF.seq	KNNIKYSLYSLKDLAEMENYEEENTNDTYENKNFSLISNS	1531
ste6YEAS.seq	RKNILLGS-----TDSVRNADCSTNENRHLIK-----	1429
msbaECOL.seq	-----	697
hlybPROV.seq	PAVIKY-----RKIF-----	463

```

-----AVKEANAYDFIMKLPDK-----
          610          620          630          640
mdr1FUGU.seq -----ATKESNAYDFIMNLPDK 1417
mdr2FUGU.seq -----ATKESNAYDFIMNLPDK 1426
mdr1CRIG.seq -----AVKEANAYDFIMKLPDK 1546
mdr1HUMA.seq -----AVKEANAYDFIMKLPDK 1555
mdr3MOUS.seq -----AVKEANAYDFIMKLPDK 1543
mdr2-RAT.seq -----AVKEANAYDFIMKLPDK 1552
mdr1CAEE.seq -----ACKMANAEKFIKTLPNG 1627
mdr4DROM.seq -----AARAANCHDFITRLPKG 1585
pmd1SCHP.seq -----DAAKLANAYDFIMTLPEQ 1666
mdr1LEIE.seq -----EACRQANIHDITMALPDR 1552
mdrPLAFF.seq MTSNELLEMKKEYQTIKDSDVVDVSKKVLIHDFVSSLPDK 1651
ste6YEAS.seq -----DACQMALLDRFILDLPDG 1480
msbaECOL.seq ----- 697
hlybPROV.seq ----- 463

```

```

FDTLVGDRGAQLSGGQKQRIAIARALVRNPKILLLDEATS
          650          660          670          680
mdr1FUGU.seq FETLVGDRGTQLSGGQKQRIAIARALVRNPKILLLDEATS 1537
mdr2FUGU.seq FETLVGDRGTQLSGGQKQRIAIARALVRNPKILLLDEATS 1546
mdr1CRIG.seq FDTLVGERGAQLSGGQKQRIAIARALVRNPKILLLDEATS 1666
mdr1HUMA.seq FDTLVGERGAQLSGGQKQRIAIARALVRNPKILLLDEATS 1675
mdr3MOUS.seq FDTLVGERGAQLSGGQKQRIAIARALVRNPKILLLDEATS 1663
mdr2-RAT.seq FDTLVGDRGAQLSGGQKQRIAIARALVRNPKILLLDEATS 1672
mdr1CAEE.seq YNTLVGDRGTQLSGGQKQRIAIARALVRNPKILLLDEATS 1747
mdr4DROM.seq YDTQVGEKGAQISGGQKQRIAIARALVRQPQVLLLDEATS 1705
pmd1SCHP.seq FSTNVGQRGFLMSGGQKQRIAIARAVISDPKILLLDEATS 1786
mdr1LEIE.seq YDTPVGPVGSLLSGGQKQRIAIARALVKRPPILLLDEATS 1672
mdrPLAFF.seq YDTLVGSNASKLSGGQKQRISIAARAIMRNPKILILDEATS 1771
ste6YEAS.seq LETLIGTGGVTLSSGGQQQRVAIARAFIRDTPILFLDEAVS 1600
msbaECOL.seq ----- 697
hlybPROV.seq ----- 463

```

```

ALDTESEAVVQAALDK--AREGRTTIVIAHRLSTIRNADV
          690          700          710          720
mdr1FUGU.seq ALDAESETIVQAALDK--VRLGRTTIVIAHRLSTIRNADI 1651
mdr2FUGU.seq ALDAESETIVQAALDK--VRLGRTTIVVAHRLSTIRNADI 1660
mdr1CRIG.seq ALDTESEAVVQAALDK--AREGRTTIVIAHRLSTVRNADI 1780
mdr1HUMA.seq ALDTESEAVVQVALDK--ARKGRTTIVIAHRLSTVRNADV 1789
mdr3MOUS.seq ALDTESEAVVQAALDK--AREGRTTIVIAHRLSTVRNADV 1777
mdr2-RAT.seq ALDTESEAEVQAALDK--AREGRTTIVIAHRLSTVRNADV 1786
mdr1CAEE.seq ALDAESEGIVQAALDK--AAKGRTTIIIAHRLSTIRNADL 1861
mdr4DROM.seq ALDPTSEKRVQSALEL--ASQGPTTLVVAHRLSTITNADK 1819
pmd1SCHP.seq ALDSKSEVLVQKALDN--ASRSRTTIVIAHRLSTIRNADN 1900
mdr1LEIE.seq ALDRKSEMEVQAALDQLIQRGGTTVVVIAHRLATIRDMDR 1792
mdrPLAFF.seq SLDNKSEYLVQKTINNLLKGNENRITIIIAHRLSTIRYANT 1891
ste6YEAS.seq ALDIVHRNLLMKAIRHW--RKGKTTIILTHELSQIESDDY 1714
msbaECOL.seq ----- 697
hlybPROV.seq -----IETLIVS----- 484

```

	I - - - - - AGFSDGVIVEQGTHSELM - EEKGIYFKLVMTQTA	
	730 740 750 760	
mdr1FUGU.seq	I - - - - - AGFSNGEIVEQGTHSQLM - EIKGVYHGLVMTQSF	1753
mdr2FUGU.seq	I - - - - - AGFSNGKIVEQGTHSQLM - EIKGVYHGLVMTQTF	1762
mdr1CRIG.seq	I - - - - - AGFDGGVIVEQGNHEELM - REKGIYFKLVMTQTA	1882
mdr1HUMA.seq	I - - - - - AGFDDGVIVEKGNHDELM - KEKGIYFKLVMTQTA	1891
mdr3MOUS.seq	I - - - - - AGFDGGVIVEQGNHDELM - REKGIYFKLVMTQTA	1879
mdr2-RAT.seq	I - - - - - AGFEDGVIVEQGSHELI - KKEGIYFRLVNMQTS	1888
mdr1CAEE.seq	I - - - - - ISCKNGQVVEVGDHRALM - AQQGLYYDLVTAQTF	1963
mdr4DROM.seq	I - - - - - VFLKDGVAEQGTHEELM - ERRGLYCELVSITQR	1921
pmd1SCHP.seq	I - - - - - VVVNAGKIVEQGSHNELL - DLNGAYARLVEAQKL	2002
mdr1LEIE.seq	IYYVKHDGAEGSRITESGTFDELL - ELDGEFAAVAKMQGV	1909
mdr1PLAFF.seq	IFVLSNRERSDNN - NNNNDDNNNNNNNNNNNNKINNEGSY	2005
ste6YEAS.seq	LYLMK - - - - - EGEVVESGTQSELLADPTTTFSTWYHLQND	1819
msbaECOL.seq	- - - - -	697
hlybPROV.seq	- - - - - VFLQLFAL - - - - -	508

	GN - - - - - EVELENEA - - - - - SESK - - E - E	
	770 780 790 800	
mdr1FUGU.seq	QK - - - - - LEDLESD - - - - - YEPW - VA - E	1804
mdr2FUGU.seq	HN - - - - - VEEENTAM - - - - - SELS - AG - E	1813
mdr1CRIG.seq	GN - - - - - EIELGNEV - - - - - GESK - NEID	1936
mdr1HUMA.seq	GN - - - - - EVELENEA - - - - - DESK - SEID	1945
mdr3MOUS.seq	GN - - - - - EIELGNEA - - - - - CKSK - DEID	1933
mdr2-RAT.seq	GS - - - - - QILSEEF - - - - - VELS - DEKA	1942
mdr1CAEE.seq	TD - - - - - AVDSAAEGK - - - - - FSRENSVAR - QTSE	2035
mdr4DROM.seq	KE - - - - - ATEADEGA - - - - - VAGR - PLQK	1975
pmd1SCHP.seq	SGGEKDQEMVEEELEDAPREIP - ITSGDDDEDNDMASLE	2119
mdr1LEIE.seq	LAGD - - - - -	1924
mdr1PLAFF.seq	IIEQGTHDSL MKNKNGIYHLMINNQKISSNKSSNNGNDNG	2125
ste6YEAS.seq	YSDAKT - - - - - IVDTETEESIHTEVESFNSQLETPKLG - - -	1921
msbaECOL.seq	- - - - -	697
hlybPROV.seq	- - - - - ITPL - - - - -	520

	- - - MSSNDS - SSLIRRRSTRGSL - - - - - T	
	810 820 830 840	
mdr1FUGU.seq	KSQ LIESFSQSS LQRRRSTRGSL - - - - - AVSE	1888
mdr2FUGU.seq	KSPVEKTVS QSSI IRRKSTRGSSF - - - - - AASE	1897
mdr1CRIG.seq	NLDMSSKDSASSLIRRRSTRRSI - - - - - R	2008
mdr1HUMA.seq	ALEMSSNDSRSSLI RKRSTRRSV - - - - - R	2017
mdr3MOUS.seq	NLDMSSKDSGSSLI RRRSTRKSI - - - - - C	2005
mdr2-RAT.seq	AGGVAPNGWKARIFRN - STKKS L - - - - - K	2011
mdr1CAEE.seq	HEGLSRQASEMDDIMNRVRSSTIG - - - - - SITN	2119
mdr4DROM.seq	SQNLSD EETDDDEE EDEE EPEL - - - - - QTS G	2059
pmd1SCHP.seq	APMMSHNTD TDTLNNKLNKDNVVFEDKTLQHVASEIVPN	2239
mdr1LEIE.seq	- - - - AKSGASVRDAKKASGHLGVILDEADLAQLDEDVPR T	2029
mdr1PLAFF.seq	SDNKSSAYKDSDTGNDADNMNSLSIHENENISNNRNCKNT	2245
ste6YEAS.seq	- - - - - SCLSNLGYDETDQLS FYEAIYQK	1987
msbaECOL.seq	- - - - -	697
hlybPROV.seq	- - - - -	520

GSNAQDRKRLSTE-EAL-----DESVPV-----VSFLRI
 850 860 870 880

mdr1FUGU.seq	GTKEEKEKFECDQDNIE-----EDENVPP---VSFFKV	1978
mdr2FUGU.seq	GTKEEKT-----E-----EDEDVPD---VSFFKV	1960
mdr1CRIG.seq	GPHDQDRKRLSTK-EAL-----DEDVPP---ISFWRI	2089
mdr1HUMA.seq	GSQAQDRKRLSTK-EAL-----DESIPP---VSFWRI	2098
mdr3MOUS.seq	GPHDQDRKRLSTK-EAL-----DEDVPP---ASFWR I	2086
mdr2-RAT.seq	SSRAHQNRRLDVETNEL-----DANVPP---VSFLKV	2095
mdr1CAEE.seq	GPVIDEKEERIGKDALS-RLKQELEENNAQK---TNLFEI	2227
mdr4DROM.seq	SSRDSGFRASTRRKRRSQRRKKKKDKEVVS K---VSFTQL	2170
pmd1SCHP.seq	LPPADVGELNEEPKSKSKSKNNHEINSLTA---LWFIHS	2350
mdr1LEIE.seq	ARQ-----NVPIDELAKW-----EVKHAK---VGF LRL	2104
mdrPLAFF.seq	AENEKEEKVPPFFKRMFRRKKKAPNN-----LRI	2329
ste6YEAS.seq	RSNVRTRRVKVEEENIGYALKQQKNTESSTGPQLLSIIQI	2107
msbaECOL.seq	-----	697
hlybPROV.seq	-----F---	523

LKLNITEWPY---FVVGILCAIINGALQPAFSIIFSKVLG
 890 900 910 920

mdr1FUGU.seq	MRYNVSEWPY---ILVGTICAMINGAMQPVFSIIFTEIIM	2089
mdr2FUGU.seq	LHLNIPWPY---ILVGLICATINGAMQPVFAILFSKIIT	2071
mdr1CRIG.seq	LKLNSSEWPY---FVVGIFCAIVNGALQPAFSIIFSKVVG	2200
mdr1HUMA.seq	MKLNLTEWPY---FVVGVFCAIINGGLQPAFAIIFSKIIG	2209
mdr3MOUS.seq	LKLNSTEWPY---FVVGIFCAIINGGLQPAFSVIFSKVVG	2197
mdr2-RAT.seq	LRLNKTEWPY---FVVGITLCAIANGALQPAFSIILSEMI A	2206
mdr1CAEE.seq	LYHARPHALS---LFIGMSTATIGGFIYPTYSVFETSFMN	2338
mdr4DROM.seq	MKLNSPEWRF---IVVGGIASVMHGATFPLWGLFFGDFFG	2281
pmd1SCHP.seq	FVRTMIEIIC---LLIGILASMICGAAYPVQAAVFARFLN	2461
mdr1LEIE.seq	MRMNKDKAWA---VALGILSSVVIGSARPASSIVMGHMLR	2215
mdrPLAFF.seq	IYKEI---FSYKKDVTIIFFSILVAGGLYPVFALLYARYVS	2443
ste6YEAS.seq	IKRMIKSIRYKKILILGLLCSLIAGATNPVFSYTF SF LLE	2227
msbaECOL.seq	-----	697
hlybPROV.seq	-----FQVVM D K V L -	550

V---F-RGDD-E-VRQKSNLFSLLFLVLGVASFITFFL--
 930 940 950 960

mdr1FUGU.seq	V---F-REKDKEIIREKSSFFCILFAVMGVVTF L T M F L Q F	2197
mdr2FUGU.seq	V---F-ADPDRDSVRRKSEFISLMFVVIGCVS F V T M F L Q -	2179
mdr1CRIG.seq	V---FTRNTDDETKRHD SN L F S L L F L I L G V I S F I T F F L --	2308
mdr1HUMA.seq	V---FTRIDDPETKRQNSNLFSLLFLALGIISFITFFL--	2317
mdr3MOUS.seq	V---FTNGGPPETQRQNSNLFSLLFLILGIIISFITFFL--	2305
mdr2-RAT.seq	I---FGPGDDT-VKQQKCNMFSLVFLGLGVH S F F T F F L --	2311
mdr1CAEE.seq	V---F---AGNPADFLSQGHFWALMFLV L A A A Q G I C S F L M -	2443
mdr4DROM.seq	I---LSDGDD-DVVRAEVLKISMI F V G I G L M A G L G N M L --	2386
pmd1SCHP.seq	I---FTDLSSTD-FLHKVNVFAVYWLILAI V Q F F A Y ---	2560
mdr1LEIE.seq	VLGEYSATKDVEALRSGTNLYAPLFIVFAVANFSG-----	2323
mdrPLAFF.seq	TLFDFAN-----LEYNSNKYSIYILLIAIAMFISETLK-	2545
ste6YEAS.seq	--GIVPSTDGKTGSSHYLAKWSLLVLGVAAADGIFNFAK-	2341
msbaECOL.seq	-----	697
hlybPROV.seq	-----VHRGFSTLNIITIALAVVAIFEITLS-	628

---QGFTFGKAGEILTKRLRYMLFRSMLRQDVSWFDDPKN
 970 980 990 1000

mdr1FUGU.seq WGFQGFCSKSGEILTLNLRRLKAFISMRRQDLSWYDNPKN 2317
 mdr2FUGU.seq ---GYCFGKSGEILTLKLRRLRAFTAMMRQDLSWYDNPQN 2284
 mdr1CRIG.seq ---QGFTFGKAGEILTKRLRYMVFKSMLRQDVSWFDDPKN 2416
 mdr1HUMA.seq ---QGFTFGKAGEILTKRLRYMVFRSMLRQDVSWFDDPKN 2425
 mdr3MOUS.seq ---QGFTFGKAGEILTKRLRYMVFKSMLRQDVSWFDDPKN 2413
 mdr2-RAT.seq ---QGFTFGKAGEILTTRLRSMAFKAMLRQDMSWFDDHKN 2419
 mdr1CAEE.seq ---TFFMGIASESLTRDLRNKLFNRVLSQHIGFFDSPQN 2548
 mdr4DROM.seq ---QTYMFTTAGVKMTTRLRKRKAFGTIIGQDIAYFDDERN 2494
 pmd1SCHP.seq -AISNFAMTYAMEAVLQRIRYHLFRITLLRQDVEFFDRSEN 2674
 mdr1LEIE.seq WILHGF-YGYAGEHLTTKIRVLLFRQIMRQDINFFDIPGR 2437
 mdrPLAFF.seq ---NYYNKIGEKVEKTMKRRLFENILYQEMSFFDQDKN 2650
 ste6YEAS.seq ---GFLDCCSEYWMVMDLRNEVMEKLTRKNMDWFSGENN 2446
 msbaECOL.seq ---VSNRMR----- 715
 hlybPROV.seq -GLRTYIFTHSTSRIDVELGAKLFRHLLALPISYFESR-- 739

TVGALTTURLANDAAQVQGATGVRLAVITQNIANLGTGII
 1010 1020 1030 1040

mdr1FUGU.seq TVGALTTURLAADAHVQGAAGVRLAVMTQNFANLGTSI II 2437
 mdr2FUGU.seq TVGALTTURLAADAQVQGAAGVRLATIMQNFANLGTSI II 2404
 mdr1CRIG.seq TTGALTTURLANDAGQVKGATGARLAVITQNIANLGTGII 2536
 mdr1HUMA.seq TTGALTTURLANDAAQVKAIGSRLAVITQNIANLGTGII 2545
 mdr3MOUS.seq TTGALTTURLANDAAQVKGATGSRLAVIFQNIANLGTGII 2533
 mdr2-RAT.seq STGALSTRLATDAAQVQGATGTRLALIAQNTANLGTGII 2539
 mdr1CAEE.seq ASGKISTRLATDVPNLRTAIDFRFSTVITTLVSMVAGIGL 2668
 mdr4DROM.seq SVGALCSRLASDCSNVQGATGARVGTMLQAVATLVVGMVV 2614
 pmd1SCHP.seq TVGAIITTSLSSTKIQSLEGLSGPTLGTFFQILTNIISVTIL 2794
 mdr1LEIE.seq DAGTLAGMLSGDCEAVHQLWGPSIGLKVQTMCIASGLVV 2557
 mdrPLAFF.seq TPGVLSAHINRDVHLLKTGLVNNIVIFSHFIMFLVSMVM 2770
 ste6YEAS.seq KASEISALVLDLRLDLRSLVSEFLSAMTSFVTVSTIGLIW 2566
 msbaECOL.seq -----LQGMKMVSASSISDPIIQLIASLAL 787
 hlybPROV.seq -----RVGDTVARVRELDQIRNFLTGQALTSIL-DLLF 832

SFIY---GW---QLTLLLLAIVPIIAIAGAVEMKMLSGQ
 1050 1060 1070 1080

mdr1FUGU.seq SFVY---GW---ELTLLILAVVPIILAVAGAAEVKLLTGH 2536
 mdr2FUGU.seq AFVY---GW---ELTLLILAVVPLIAAAGAAEIKLLAGH 2503
 mdr1CRIG.seq SLIY---GW---QLTLLLLAIVPIIAIAGVEMKMLSGQ 2635
 mdr1HUMA.seq SFIY---GW---QLTLLLLAIVPIIAIAGVEMKMLSGQ 2644
 mdr3MOUS.seq SLIY---GW---QLTLLLLAIVPIIAIAGVEMKMLSGQ 2632
 mdr2-RAT.seq SFIY---GW---QLTLLLLSVVPIIAGVEMKMLAGN 2638
 mdr1CAEE.seq AFFY---GW---QMALLIIAILPIVAFGQYLRGRRFTGK 2767
 mdr4DROM.seq GFVF---SW---QQTLLTLVTLPLVCLSVYLEGRFIMKS 2713
 pmd1SCHP.seq SLAT---GW---KLGLVTLSTSPVIITAGYYRVRALDQV 2893
 mdr1LEIE.seq GFIY---QW---KLALVALACMPLMI--GCSLTRRLMIN 2650
 mdrPLAFF.seq SFYFCPIV---AAVLTFIYFINMRVFAVRARLTKSKEIEK 2881
 ste6YEAS.seq ALV---SGW---KLSLVCISMFLIIFSAIYGGILOKC 2665
 msbaECOL.seq AFVL----- 802
 hlybPROV.seq SFIFFAVMWYYS PKLTLVILFSLPCYATWSIFISPILRRR 952

ALKDKKEL-----EGAGKIATEAIEENIRTVVSLTR
 1090 1100 1110 1120

mdr1FUGU.seq	AAEDKKEL-----EMAGKIATEAIEENVRTVVSLTR	2626
mdr2FUGU.seq	AAKDKKEL-----EKAGKIATEAIEENVRTVVSLSR	2593
mdr1CRIG.seq	ALKDKKEL-----EGSGKIATEAIEENFRTVVSLTR	2725
mdr1HUMA.seq	ALKDKKEL-----EGAGKIATEAIEENFRTVVSLTQ	2734
mdr3MOUS.seq	ALKDKKEL-----EGSGKIATEAIEENFRTVVSLTR	2722
mdr2-RAT.seq	AKRDKKEM-----EAAGKIATEAIEENIRTVVSLTQ	2728
mdr1CAEE.seq	NVKSASEF-----EADSGKIAIEAIEENVRTVQALAR	2857
mdr4DROM.seq	AQKAKASI-----EEASQVAVEAITNIRTVNGLCL	2803
pmd1SCHP.seq	QEKLSAAY-----KESAAFACESTSAIRTVASLNR	2983
mdr1LEIE.seq	GYTKSREG-----DTDDTIVTEALS NVRTV TSLNM	2740
mdrPLAFF.seq	KENMSSGVFAFSSDDEMFKDPSFLIQEAFYNMHTVINYGL	3001
ste6YEAS.seq	ETDYKTSV-----AQLNCLYQIVTNIKTIKCLQA	2755
msbaECOL.seq	-----	802
hlybPROV.seq	LDDKFARN-----ADNQSFVLESVTAINTIKAMAV	1042

EQKFENMYAESLDVPHYRNSVKKAHIFGLTFSFTQAMMYFS
 1130 1140 1150 1160

mdr1FUGU.seq	EPTFVALYEENLTVPHYKNSQKKAKIYGLTYSFSQAMIFFV	2746
mdr2FUGU.seq	EPKFECLYEENLRVPHYKNSQKKAHVYGLTYSFSQAMIYFA	2713
mdr1CRIG.seq	EQKFENMYAQS LQIPYRNALKKAHVFGITFSFTQAMMYFS	2845
mdr1HUMA.seq	EQKFEHMYAQS LQVPYRNSLRKAHIFGITFSFTQAMMYFS	2854
mdr3MOUS.seq	EQKFETMYAQS LQIPYRNAMKKAHVFGITFFFQAMMYFS	2842
mdr2-RAT.seq	ERKFESMYVEKLVHGPYRNSVRKAHIYGITFSISQAFMYFS	2848
mdr1CAEE.seq	EDTFYENFCEKLDIPHKEAIKEAFIQGLSYGCASSVLYLL	2977
mdr4DROM.seq	ERQVLDQYVQQIDRV DVACRRKVRFRGLV FALGQAAPFLA	2923
pmd1SCHP.seq	EENVFAEYCDSLIKPGRESAIASLKSGLFFSAAQGVTF LI	3103
mdr1LEIE.seq	KEDCVEAFQAALREEAPRSVRKGI IAGGIYGITQFIFYGV	2860
mdrPLAFF.seq	EDYFCNLIIEKAIDYKKNKGQKRRRIIVNAALWGFSSAQ LFI	3121
ste6YEAS.seq	EFHFQLT YHDLKIKMQQIASKRAIATGFGISMTNMIVMCI	2875
msbaECOL.seq	-----	802
hlybPROV.seq	SPQMTNIWDKQLAGYVAAGFKVTVLATIGQQGIQLIQKAV	1162

YAACFRFGAYLVAAGL--MTFEDVLLVFS AIVFGAMAVGQ
 1170 1180 1190 1200

mdr1FUGU.seq	YAACFRFGAWLIEAGR--MDVEGVFLVVM TMLYGAMAVGE	2860
mdr2FUGU.seq	YAACFRFGAWLIEAGR--MDVEGVFLVVS AVLYGAMAVGE	2827
mdr1CRIG.seq	YAACFRFGAYLVAREL--MTFENVLLVFS AIVFGAMAVGQ	2959
mdr1HUMA.seq	YAGCFRFGAYLV AHKL--MSFEDVLLVFS AVVFGAMAVGQ	2968
mdr3MOUS.seq	YAACFRFGAYLV TQQL--MTFENVLLVFS AIVFGAMAVGQ	2956
mdr2-RAT.seq	YAGCFRFGSYLIVNGH--MRFKDVILVFS AIVLGAVLGH	2962
mdr1CAEE.seq	NTCAYRMGLALIITDPPTMQPMRVL RVMYAITISTSTLGF	3097
mdr4DROM.seq	YGISMYYGILVAEER--MNYEDIKVAEALIFGSWMLGQ	3037
pmd1SCHP.seq	NALTFWYGSTLMRKG--EYNIVQFYTCFIAIVFGIQQAGQ	3217
mdr1LEIE.seq	YALCFWYGSKLIDKGEAE--FKDVMIASMSILFGAQNAGE	2974
mdrPLAFF.seq	NSFAYWFGSFLIKRG--TILVDDFMKSLFTFI FTGSYAGK	3235
ste6YEAS.seq	QAI IYYYGLKLVMM--IHEYTSKEMFTTFTLL LFTIMSCTS	2989
msbaECOL.seq	YAASF-----PSVMDSL TAGTITVVFSSMIALMRPLKS	898
hlybPROV.seq	MIINLWLGAHLV ISGDLSIGQLIAFNMLAGQIVA--PVIR	1276

ASSFAPDYAKAKASASHLFLMLIEKKPEIDSYSTEG-LTP-
 1210 1220 1230 1240

mdr1FUGU.seq	ANTYAPNFAKAKISASHLTM LINRQPAIDNLSEEE-ARL-	2977
mdr2FUGU.seq	ANTFAPNYAKAKMAASYLMMLINKKPAIDNLSEEG-TSP-	2944
mdr1CRIG.seq	VSSFAPDYAKAKVSAASHIIMIIEKVPSIDSYSTGG-LKP-	3076
mdr1HUMA.seq	VSSFAPDYAKAKISAAHIIMIIEKTPLIDSYSTEG-LMP-	3085
mdr3MOUS.seq	VSSFAPDYAKATVSASHIIRIIEKTPEIDSYSTQG-LKP-	3073
mdr2-RAT.seq	ASSFAPDYAKAKLSAAYLFSLFRQPLIDSYREG-MWP-	3079
mdr1CAEE.seq	ATSYFPEYAKATFAGGIIFGMLRKRISKIDSLSLAG-EK--	3211
mdr4DROM.seq	ALAYAPNVNDAILSAGRLMDLFRKRTSTQPNPPQSPYNTV-	3157
pmd1SCHP.seq	FFGYSADVTKAKAAAGEIKYLSESKPKIDTWSTEGKKVE-	3337
mdr1LEIE.seq	AGAFATKLADAEASAKRVFSVIDRVPDVIDEQAGNKDLG-	3094
mdrPLAFF.seq	LMSLKGDS ENAKLSFEKYYP LMIRKSNIDVRDDG-----	3340
ste6YEAS.seq	LVSQIPDISRGQRAASWIYRILDEKHNTLEVENNNARTVG	3109
msbaECOL.seq	LTNVNAQFQRGMAACQTLFTILDSEQEKD-----EGKRV I-	1006
hlybPROV.seq	LAQLWQDFQQVGVISVTRLGDV LNYPT E-----SYQGKLT L-	1384

---EKLEGN-----VEFRNVVFNYPTRPDIPV LQGLNLEV
 1250 1260 1270 1280

mdr1FUGU.seq	---EKYDGN-----VLFEDVKFNYP SRPDVPV LQGLNLEV	3070
mdr2FUGU.seq	---EKYDGN-----VHFEGVKFNYP SRPDVTI LQGLNLKV	3037
mdr1CRIG.seq	---NTLEGN-----VKFNEVVFNYPTRPDIPV LQGLNLEV	3169
mdr1HUMA.seq	---NTLEGN-----VTFGEVVFNYPTRPDIPV LQGLSLEV	3178
mdr3MOUS.seq	---NMLEGN-----VQFSGFVFNYPTRPSIPV LQGLSLEV	3166
mdr2-RAT.seq	---DKFEGS-----VTFNEVVFNYPTRANVPV LQGLSLEV	3172
mdr1CAEE.seq	---KKLYGK-----VIFKNVRFAYPERPEIEILKGLSFSV	3304
mdr4DROM.seq	---EKSEGD-----IVYENVGFEYPTRKGTPI LQGLNLTI	3250
pmd1SCHP.seq	---SLQSA-----IEFRQVEFSYPTRRHIKVL RGLNLTV	3430
mdr1LEIE.seq	---E--GCD-----IEYRNVQFIYSARPKQV V LASVNMRF	3181
mdrPLAFF.seq	--GIRINKNLIK GKVDIKDVNFRYISRPNVPIYKNLSFTC	3451
ste6YEAS.seq	IAGHTYHGKEKKPIVSIQNLTFAYPSAPTAFVYKNMNFDM	3229
msbaECOL.seq	---ERATGD-----VEFRNVTF TY PGR-DVPALRNINLKI	1096
hlybPROV.seq	---PEINGD-----ISFRNIRFRYKPD API-ILNNINLNI	1474

KKGQTLALVGSSGCGKSTVVQLLERFYD-----
 1290 1300 1310 1320

mdr1FUGU.seq	QKGETLALVGSSGCGKSTTIQLLERFYD-----	3157
mdr2FUGU.seq	KKGETLALVGSSGCGKSTTIQLLERFYD-----	3124
mdr1CRIG.seq	KKGQTLALVGSSGCGKSTVVQLLERFYD-----	3256
mdr1HUMA.seq	KKGQTLALVGSSGCGKSTVVQLLERFYD-----	3265
mdr3MOUS.seq	KKGQTLALVGSSGCGKSTVVQLLERFYD-----	3253
mdr2-RAT.seq	KKGQTLALVGSSGCGKSTVVQLLERFYD-----	3259
mdr1CAEE.seq	EPGQTLALVGPSGCGKSTVVALLERFYD-----	3391
mdr4DROM.seq	KKSTTVALVGPSGSGKSTCVQLLR FYD-----	3337
pmd1SCHP.seq	KPGQFVAFVGS SGC GKSTTIGLIERFYD-----	3517
mdr1LEIE.seq	GDATSNGLIGQTGCGKSTVIQMLARFYE-----	3268
mdrPLAFF.seq	DSKKT TAIVGETSGSGKSTFMNLLRFYDLKNDHIILKNDM	3571
ste6YEAS.seq	FCGQTLGIIIGESGTGKSTLVLLLT KLYN-----	3316
msbaECOL.seq	PAGKTVALVGRSGSGKSTIASLITRFYD-----	1183
hlybPROV.seq	KQGEIIGIVGRSGSGKSTLT KLIQRFYI-----	1561

	----- PRAGE -----	
	1330 1340 1350 1360	
mdr1FUGU.seq	----- PREGR	3169
mdr2FUGU.seq	----- PREGR	3136
mdr1CRIG.seq	----- PMAGT	3268
mdr1HUMA.seq	----- PLAGK	3277
mdr3MOUS.seq	----- PMAGS	3265
mdr2-RAT.seq	----- PMAGT	3271
mdr1CAEE.seq	----- TLGGE	3403
mdr4DROM.seq	----- PVS GS	3349
pmd1SCHP.seq	----- CDNGA	3529
mdr1LEIE.seq	----- RRSGL	3280
mdrPLAFF.seq	TNFQDYQNNNNNSLVLKNVNEFSNQSGSAEDYTVFNNSG	3691
ste6YEAS.seq	----- CEVGK	3328
msbaECOL.seq	----- IDEGE	1195
hlybPROV.seq	----- PENGQ	1573

	VLLDGV DVKQLNVQWLR AQIGIVSQEPVLFDCSIAENIAY	
	1370 1380 1390 1400	
mdr1FUGU.seq	VLLDGV DVKQLNVHWR S Q I G I V S Q E P V L F D C S L A E N I A Y	3289
mdr2FUGU.seq	VSLDGV NVKQLNIHWL R S Q I G I V S Q E P V L F D C S L A E N I A Y	3256
mdr1CRIG.seq	VFLDGKEV NQLNVQWLR AHLG I V S Q E P I L F D C S I A E N I A Y	3388
mdr1HUMA.seq	VLLDGKEIKRLNVQWLR AHLG I V S Q E P I L F D C S I A E N I A Y	3397
mdr3MOUS.seq	VFLDGKEIKQLNVQWLR AQLG I V S Q E P I L F D C S I A E N I A Y	3385
mdr2-RAT.seq	VLLDGQEAKKLN VQWLR AQLG I V S Q E P I L F D C S I A K N I A Y	3391
mdr1CAEE.seq	IFIDGSEIKTLNPEHTRS QIA I V S Q E P T L F D C S I A E N I I Y	3523
mdr4DROM.seq	VNLSGVPSTEFPLD TLR SKLGLV S Q E P V L F D R T I A E N I A Y	3469
pmd1SCHP.seq	VLV DGVNVRDYNINDYR KQIALV S Q E P T L Y Q G T V R E N I V L	3649
mdr1LEIE.seq	ISVNGRDLSSLDIAEWR RN I S I V L Q E P N L F S G T V R E N I R Y	3400
mdrPLAFF.seq	ILLDDINICDYNLRDLRNLFS I V S Q E P M L F N M S I Y E N I K F	3811
ste6YEAS.seq	IKIDGTDVNDWNLTSLRKEISVVEQKPLL FNGTIRDNLTY	3448
msbaECOL.seq	ILMDGHD LREYTLASLRN QVALV SQNVHLFN DTVANNIAY	1315
hlybPROV.seq	VLIDGHDLALADPNWLR RQVGVV LQDNVLLNRSI IDNIAL	1693

	GDNSRV-VSMEEIVEAAKEANI H D F I E S L P Q G Y D T R V G D K	
	1410 1420 1430 1440	
mdr1FUGU.seq	GDNSRS-VSMDEIVAAKAANIHSFIEGLPQRYDTQAGDK	3406
mdr2FUGU.seq	GDNSRS-VSMDEIVLAAKAANIHSFIEGLPQRYDTQAGDK	3373
mdr1CRIG.seq	GDNSRV-VSQDEIERAAKEANIHQFIESLPDKYNTRVGDK	3505
mdr1HUMA.seq	GDNSRV-VSQEEIVRAAKEANIHFIESLPNKYSTKVGDK	3514
mdr3MOUS.seq	GDNSRV-VSYEEIVRAAKEANIHQFIDS LPDKYNTRVGDK	3502
mdr2-RAT.seq	GDNSRV-VSQDEIVRAAKEANIHPFIETLPQKYETRVGDK	3508
mdr1CAEE.seq	GLDPSS-VTMAQVEEAARLANIHNFI AELPEGFETRVGDR	3640
mdr4DROM.seq	GNNFRDDVSMQEIIEAAKKSNIHNFI SALPQGYDTRLG-K	3586
pmd1SCHP.seq	G--ASKDVSEEEIEACKKANIHEFILGLPNGYNTLCGQK	3763
mdr1LEIE.seq	A---REGATDEEVEEAARLAHIHHEI IKWTDGYDTEVGYK	3511
mdrPLAFF.seq	G---REDATLEDVKRVSKFAAIDEFIESLPNKYDTNVPY	3922
ste6YEAS.seq	GLQ--DEILEIEMYDALKYVGIHDFV ISSPQGLDTRI--D	3556
msbaECOL.seq	A--RTEQYSREQIEEAARMAYAMDFINKMDNGLD TVIGEN	1429
hlybPROV.seq	AD---PGMPVEKVIHAAKLAGAHDFISELREGYNTIVGEQ	1804

GTQLSGGQKQRIAIARALVRNPKILLLDEATSALDTESEK
 1450 1460 1470 1480

mdr1FUGU.seq	GTQLSGGQKQRVAIARAIIRNPKLLLLLDEATSALDTESEK	3526
mdr2FUGU.seq	GTQLSGGQKQRVAIARAIIRNPKLLLLLDEATSALDTESEK	3493
mdr1CRIG.seq	GTQLSGGQKQRIAIARALVRQPHILLLDEATSALDTESEK	3625
mdr1HUMA.seq	GTQLSGGQKQRIAIARALVRQPHILLLDEATSALDTESEK	3634
mdr3MOUS.seq	GTQLSGGQKQRIAIARALVRQPHILLLDEATSALDTESEK	3622
mdr2-RAT.seq	GTQLSGGQKQRIAIARALIRQPRVLLLDEATSALDTESEK	3628
mdr1CAEE.seq	GTQLSGGQKQRIAIARALVRNPKILLLDEATSALDTESEK	3760
mdr4DROM.seq	TSQLSGGQKQRIAIARALVRNPKILLLDEATSALDLESEK	3706
pmd1SCHP.seq	GSSLSGGQKQRIAIARALIRNPKILLLDEATSALDSHSEK	3883
mdr1LEIE.seq	GRALSGGQKQRIAIARGLLRRPRLLLLDEATSALDSVTEA	3631
mdrPLAFF.seq	GKSLSGGQKQRIAIARALLREPKILLLDEATSSLDNSSEK	4042
ste6YEAS.seq	TTLSSGGQAQRLCIARALLRKSILILDECTSALDSVSSS	3676
msbaECOL.seq	GVLSSGGQRQRIAIARALLRDSPIILILDEATSALDTESEK	1549
hlybPROV.seq	GAGLSSGGQRQRIAIARALVNNPKILIFDEATSALDYESEH	1924

VVQEALD--KAREGRTCIVIAHRLSTIQNADLIVVFQNGV
 1490 1500 1510 1520

mdr1FUGU.seq	VVQEALD--QARKGRTCIVVAHRLSTIQNADCIAVFQGGV	3640
mdr2FUGU.seq	VVQEALD--QARKGRTCIVVAHRLSTIQNADCIAVFQGGV	3607
mdr1CRIG.seq	VVQEALD--KAREGRTCIVIAHRLSTIQNADLIVVIQNGK	3739
mdr1HUMA.seq	VVQEALD--KAREGRTCIVIAHRLSTIQNADLIVVFQNGR	3748
mdr3MOUS.seq	VVQEALD--KAREGRTCIVIAHRLSTIQNADLIVVIQNGK	3736
mdr2-RAT.seq	VVQEALD--KAREGRTCIVIAHRLSTIQNADLIVVIDNGK	3742
mdr1CAEE.seq	VVQEALD--RAREGRTCIVIAHRLNTVMNADCIAVVSNGT	3874
mdr4DROM.seq	VVQQAALD--EARSGRCTLTIAHRLTTVRNADLICVLRKRV	3820
pmd1SCHP.seq	VVQEALN--AASQGRTTVAIAHRLSSIQDADCIFVFDGGV	3997
mdr1LEIE.seq	KVQEGIEAFQAKYKVTTSIAHRLTTIRHCDQIILLDSGC	3751
mdrPLAFF.seq	LIEKTIVDIKDKADKTIITIAHRIASIKRSDKIVVFNNPD	4162
ste6YEAS.seq	IINEIVK--KGPPALLTMVITHSEQMMRSCNSIAVLKDGK	3790
msbaECOL.seq	AIQAALDELQ--KNRTSLVIAHRLSTIEKADEIVVVEDGV	1663
hlybPROV.seq	VIMRNMHKI--CQGRTVIIIAHRLSTVKNADRIIVMEKQG	2038

-----VVEHGTHQQLLAQK-GIYFSLVSVQAG--SS-
 1530 1540 1550

mdr1FUGU.seq	-----VVEKGTHTQQLIAKK-GVYHMLVTKQMGYHS-G	3730
mdr2FUGU.seq	-----VVEKGTHTQQLIAKK-GVYHMLVTKQMGYHN-D	3697
mdr1CRIG.seq	-----VKEHGTHQQLLAQK-GIYFSMVS SVQAGAKR.	3829
mdr1HUMA.seq	-----VKEHGTHQQLLAQK-GIYFSMVS SVQAGTKRQ.	3841
mdr3MOUS.seq	-----VKEHGTHQQLLAQK-GIYFSMVS SVQAGAKRS.	3829
mdr2-RAT.seq	-----VKEHGTHQQLLAQK-GIYFSMVNIQAGTQNL.	3835
mdr1CAEE.seq	-----IIEKGTHTQLMSEK-GAYYKLTQKQM-TEKK.	3964
mdr4DROM.seq	-----VVEHGTHDELMLALN-KIYANLYLMQQV--SG.	3907
pmd1SCHP.seq	-----TCEAGTHAELVKQR-GRYYELV-VEQGLNKA.	4087
mdr1LEIE.seq	-----IIEQGSHEELMALG-GEYKTRYDLYMSALS.	3841
mdrPLAFF.seq	RNGTFVQSHGTHDELLSAQDGIYKKYVKLA-----K.	4258
ste6YEAS.seq	-----VVERGNFDTLYNNRGE L FQIV-----SNQSS.	3871
msbaECOL.seq	-----IVERGTHNDLLEHR-GVYAQLHKMQFG---Q.	1747
hlybPROV.seq	-----IIEQGKHKELLSDPESLYHYLHQLQ-----SD.	2122

Percent Similarity

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
1	100.0	79.9	60.7	60.2	59.1	59.2	41.2	39.9	37.5	30.9	19.7	23.1	28.1	22.2	1
2	21.7	100.0	63.3	63.6	62.6	62.5	43.3	41.8	38.9	32.4	22.1	23.3	29.7	23.0	2
3	51.0	48.0	100.0	86.7	92.4	71.7	42.8	39.8	38.1	32.3	24.8	23.9	29.2	22.9	3
4	51.2	47.9	14.1	100.0	86.7	73.1	43.1	39.9	37.2	32.6	22.9	23.2	30.2	23.2	4
5	53.6	49.2	7.6	14.4	100.0	71.2	42.8	39.3	37.9	31.7	24.2	23.3	29.0	23.7	5
6	54.9	49.9	33.2	30.7	33.6	100.0	41.3	39.2	37.3	32.5	21.8	22.5	30.4	22.7	6
7	100.0	96.5	95.9	94.9	95.2	100.0	100.0	32.2	31.7	28.9	21.1	23.6	29.8	23.2	7
8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	32.3	32.3	29.4	18.6	21.0	28.3	22.2	8
9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	31.0	20.0	23.0	28.3	22.3	9
10	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	19.6	19.0	25.9	20.2	10
11	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	17.0	23.3	16.8	11
12	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	25.0	18.9	12
13	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	27.6	13
14	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	14
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	

mdr1FUGU.seq

mdr2FUGU.seq

mdr1CRIG.seq

mdr1HUMA.seq

mdr3MOUS.seq

mdr2-RAT.seq

mdr1CAEE.seq

mdr4DROM.seq

pmd1SCHP.seq

mdr1LEIE.seq

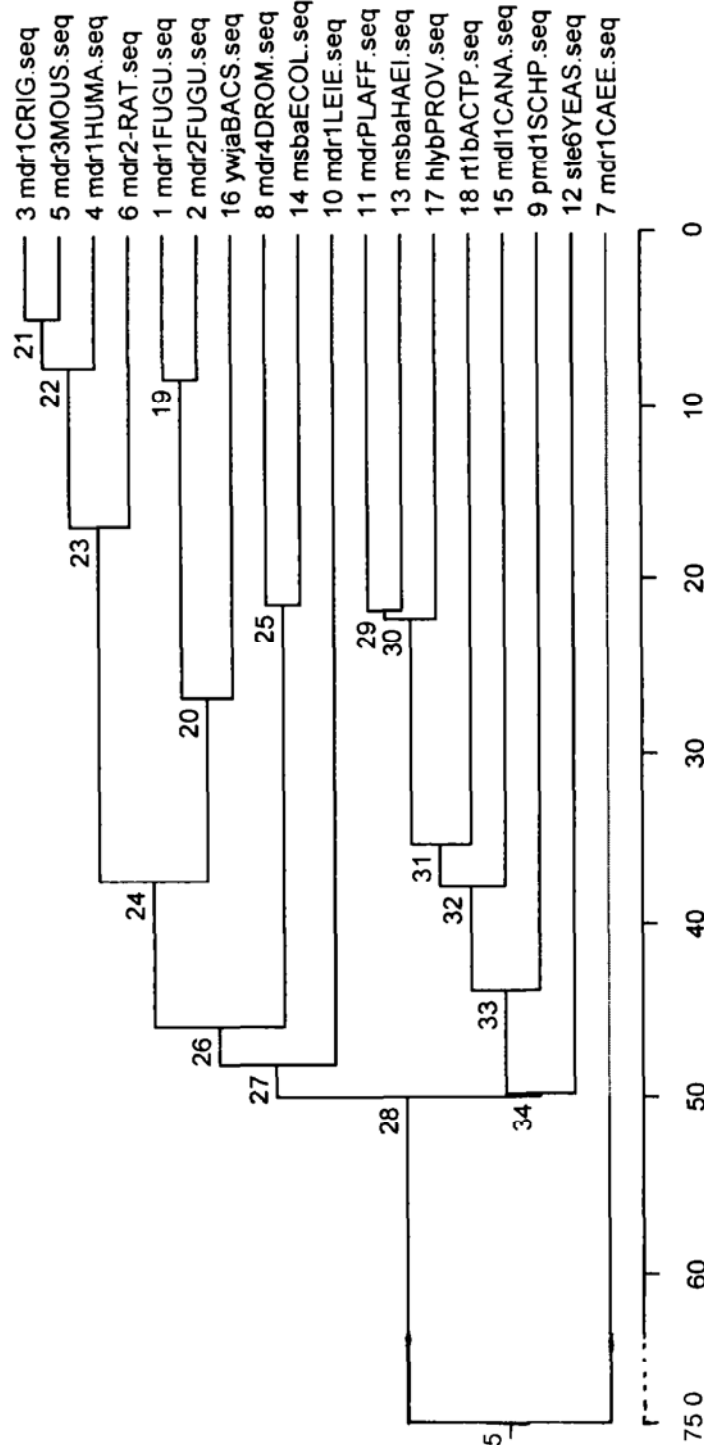
mdrPLAFF.seq

ste6YEAS.seq

msbaECOL.seq

hlybPROV.seq

它们彼此之间的进化距离如下图所示:



- 利用 Dnastar 软件对 fuguMDR1 cDNA 和 fuguMDR2 cDNA 核苷酸序列与人的 MDR1 cDNA 核苷酸序列进行比较, 浅色条带表示三者有共同的碱基序列:
fuguMDR1 cDNA 和 fuguMDR2 cDNA 与人 MDR1 cDNA 核苷酸序列比较

	TTCCGXTTTGCAGATXGXTGGGACATCTTGXTGXTCATCA	
	10 20 30 40	
mdr1FUGU.seq	TTCCGGTTTGCAGATGGTTGGGACATCTTGATGGTCACCA	40
mdr2FUGU.seq	TTCCGATTTGCAGACCGCTGGGACGTCGTGCTGCTCATCA	40
MDR1HUMA.SEQ	TTTCGCTATTCAAATTGGCTTGACAAGTTGTATATGGTGG	40
	TCGGAACXXTGATGGCCATXGTCAACGGGXCXGTGATTCC	
	50 60 70 80	
mdr1FUGU.seq	TCGGAGTGCTGATGGCCATTGTGAACGGGCTGGTGAATCC	80
mdr2FUGU.seq	GCGGGACAGTGATGGCCATGGTCAACGGCACCGTGATGCC	80
MDR1HUMA.SEQ	TGGGAAC TTTGGCTGCCATCATCCATGGGGCTGGACTTCC	80
	TCTCATGTGXATXGTGTTTGGAGAAATGACXGACAGCTTT	
	90 100 110 120	
mdr1FUGU.seq	TCTGATGTGTATCGTGTTTGGTGAGATGACTGACAGCTTC	120
mdr2FUGU.seq	CCTCATGTGCATTGTCTTTGGAGAAATGACGGACAGTTTT	120
MDR1HUMA.SEQ	TCTCATGATGCTGGTGT TTTGGAGAAATGACAGATATCTTT	120
	AT --- AXAXG CXGAAATGTCCCAAXAXCACAAACXCAAGCA	
	130 140 150 160	
mdr1FUGU.seq	ATCCAGGAAGCCAAACTGTCCCAAAACCAACACAAGCA	160
mdr2FUGU.seq	AT --- ATACGCTGACATGGCCCAA --- CACAACGCAAGTG	154
MDR1HUMA.SEQ	GC --- AAATGCAGGAAATTTAGAAGATCTGATGTCAAACA	157
	XCXCAATTCAACTAXTACTAATGTTAXAAGXXTGTTTCX -	
	170 180 190 200	
mdr1FUGU.seq	ACCCCACTTCAACAATGTCTAATGTTTAAAGAGTGTGCT -	200
mdr2FUGU.seq	GCTGGAATTCTACTACTACTATTCTGAACAGCACGTTAC -	194
MDR1HUMA.SEQ	TCACTAATAGAAGTGATATCAATGATACAGGGTTCTTCAT	197
	-----AGGAXGACATGCA XAGATTGCGCCATTTATTAC	
	210 220 230 240	
mdr1FUGU.seq	-----GCATCATTGCGCATCTATTCTCCATCTACTAC	231
mdr2FUGU.seq	-----AGGAGGACATGCAAAGATTGCGCCATTTATTAC	225
MDR1HUMA.SEQ	GAATCTGGAGGAAGACATGACCAGGTATGCCTATTATTAC	237
	TCCGTCTTGGGXTTTGTGCTGGTXGCXGCXTACATGC	
	250 260 270 280	
mdr1FUGU.seq	TCCATCTTGGGGTTTGTGCTGTGCTGGTAGTGGCGTACCTGC	271
mdr2FUGU.seq	TCCGTCTTGGGATTTGTTGTGCTGCTGGCCGCCTACATGC	265
MDR1HUMA.SEQ	AGTGGAATTGGTGCTGGGGTGCTGGTTGCTGCTTACATTC	277
	AGGTGTCXTTTGTGGACCCTAACAGCCGGGCGXCAGGTXAA	
	290 300 310 320	
mdr1FUGU.seq	AGATGTCTCTGTGGACCCTAACGGCCGCGCGGCAGGCCAA	311
mdr2FUGU.seq	AGGTGTCCTTCTGGACCATAACAGCCGGGCGCCAGGTGAA	305
MDR1HUMA.SEQ	AGGTTTCATTTTGGTGCTGGCAGCTGGAAGACAAATACA	317

ACGAATTCGCAAXTTGTTTTTCCACGGCATCATGCAGCAG
 330 340 350 360

mdr1FUGU.seq ACGAATTCGCGAGTTGTTTTTCCACGGCATCATGCAGCAG 351
 mdr2FUGU.seq ACGCATCCGCAGCTTGTTTTTTCCACTGCATCATGCAGCAG 345
 MDR1HUMA.SEQ CAAAATTAGAAAACAGTTTTTTTCATGCTATAAATGCGACAG 357

GAGATCAGCTGGTTTTGACGTGAACGAXACAGGGGAGCTCA
 370 380 390 400

mdr1FUGU.seq GACATCAGCTGGTATGACGTGACTGAGACAGGAGAGCTCA 391
 mdr2FUGU.seq GAGATCAGCTGGTTTTGACGTGAACGACACAGGGGAGCTCA 385
 MDR1HUMA.SEQ GAGATAGGCTGGTTTTGATGTGCACGATGTTGGGGAGCTTA 397

ACACXCGACTXACAGAXGAXGTCTXXAAGATXXAGGAGGG
 410 420 430 440

mdr1FUGU.seq ACACGCGTCTCACAGA-----GTGGGAGGG 416
 mdr2FUGU.seq ACACTCGACTGACGGACGACGTCTACAAGATCCAGGAGGG 425
 MDR1HUMA.SEQ ACACCCGACTTACAGATGATGTCTCTAAGATTAATGAAGT 437

CATCGGTGACAAGGTGGGXCTGCTGATCCAGGCGXXCXCC
 450 460 470 480

mdr1FUGU.seq CATCGGTGACAAGGCGGGTCTGCTGATCCAGGCGGCCTCC 456
 mdr2FUGU.seq CATCGGTGACAAGGTGGGGCTGCTGATCCAGGCGTACACC 465
 MDR1HUMA.SEQ TATTGGTGACAAAATTGGAATGTTCTTTCAGTCAATGGCA 477

ACCTTTATCACTGCCTTTATXATXGGXTTTACACXTGGXT
 490 500 510 520

mdr1FUGU.seq ACCTTTATCACTTCCTTTGTATTGGTTTTGTACATGGAT 496
 mdr2FUGU.seq ACCTTCATCACGGCCTTCATCATCGGCTTCACCACGGGCT 505
 MDR1HUMA.SEQ ACATTTTTCACTGGGTTTTATAGTAGGATTTACACGTGGTT 517

GGAAGCTXACCCTGGTCATCCTGGCCATCAGCCCTGTGCT
 530 540 550 560



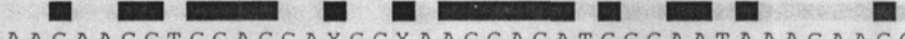
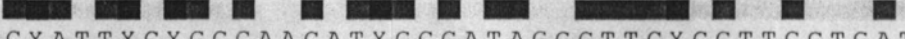
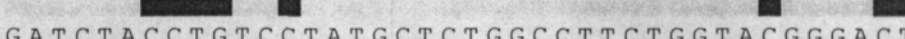
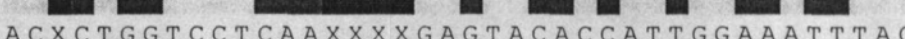
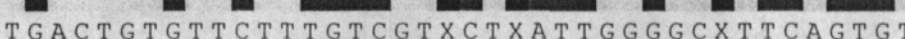
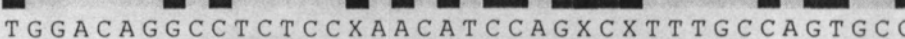
mdr1FUGU.seq GGAAGCTCACCCCTGGTCATCCTGGCCATCAGCCCTGTGTT 536
 mdr2FUGU.seq GGAAACTGACGCTGGTCATCCTGGCCGTGAGCCCCGCGCT 545
 MDR1HUMA.SEQ GGAAGCTAACCCTTGTGATTTTGGCCATCAGTCCTGTTCT 557

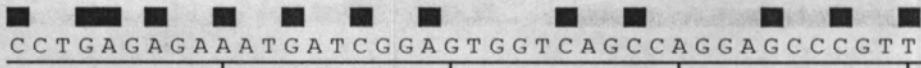
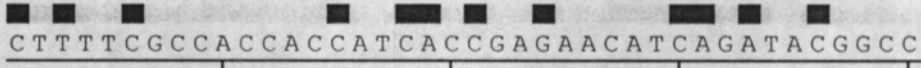
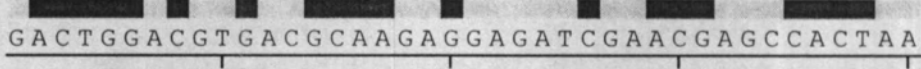
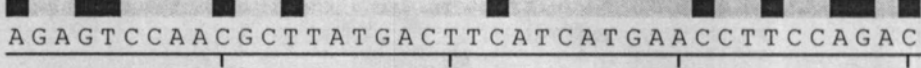
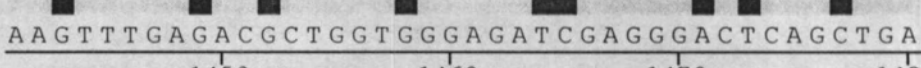
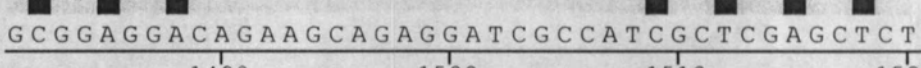
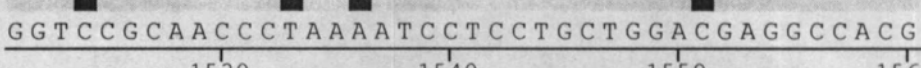
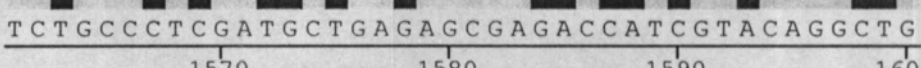
GGGXCTXTCAGCTGCCXTCTXCAGTAAGXTGCTXXCXTCC
 570 580 590 600




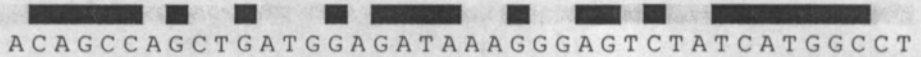
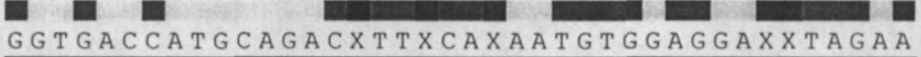
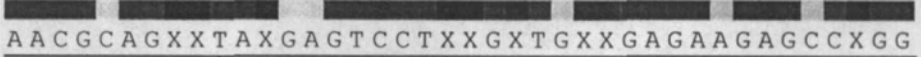
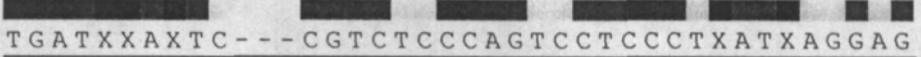
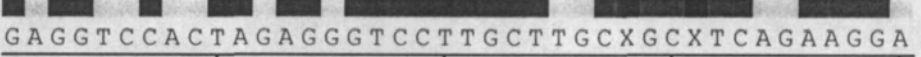
mdr1FUGU.seq GGGTCTCTCAGCTGCCCTTTACAGTAAGTTGCTGACAAGC 576
 mdr2FUGU.seq GGCCATTTTCGGCCGCCTTCTTTCAGTAAGTGCTTGCGTCC 585
 MDR1HUMA.SEQ TGGACTGTGCTGCTGTCTGGGCAAAGATACTATCTTCA 597

TTCACCAGTAAAGAGCAGACAGCGTACGCCAAAGCTGGAG
 610 620 630 640

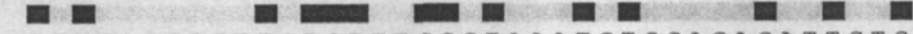
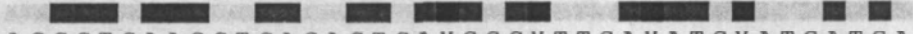
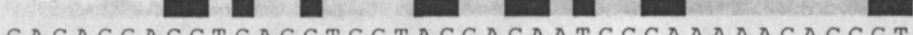
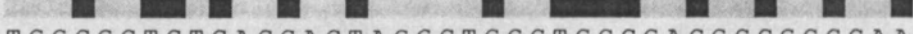
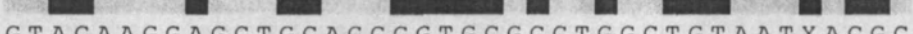
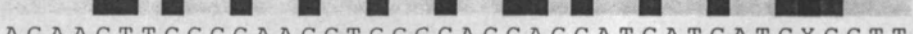
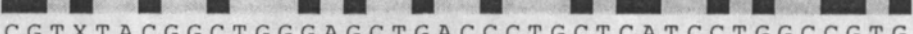
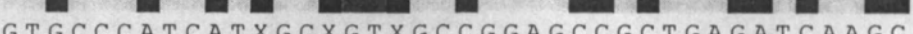
mdr1FUGU.seq TTCACCAGTAAAGAGCAGACAGCGTACGCCAAAGCTGGAG 616
 mdr2FUGU.seq TTCACCAGTAAAGAGCAGACGGCGTACGCCAAAGCCGGAG 625
 MDR1HUMA.SEQ TTTACTGATAAAGAACTCTTAGCGTATGCAAAGCTGGAG 637

		
	CXGTGGCXGAGGAGGTGCTGTCCGCCATCAGGACTGTGTT	
	650 660 670 680	
mdr1FUGU.seq	CTGTGGCAGCGGAGGTGCTATCCTCCATCAGGACTGTGTT	656
mdr2FUGU.seq	CCGTGGCGGAGGAAGTGTGTCCGCCATCAGGACCGTGTT	665
MDR1HUMA.SEQ	CAGTAGCTGAAGAGGTCTTGGCAGCAATTAGAAGTGTGAT	677
		
	TGCCTTCAGTGGXCAAAXXAAAGAXATTGAAAGATACCAC	
	690 700 710 720	
mdr1FUGU.seq	TGCCTTCAGTGGCCAAAGAAAAGCCATCAAAGATATCAT	696
mdr2FUGU.seq	CGCCTTCAGTGGTCCAGACCAGAGAGATTGAGAGATACCAC	705
MDR1HUMA.SEQ	TGCATTTGGAGGACAAAAGAAAGAACTTGAAAGGTACAAC	717
		
	AAGAACCTGGAGGAXGCXAAGGACATGGGAATAAAGAAGG	
	730 740 750 760	
mdr1FUGU.seq	AAGAACCTGGAGGATGCGAGGGACATGGGAATAAAGAAGG	736
mdr2FUGU.seq	AAGAACCTGCGGGACGCAAAGGACGTGGGAGTGAAGAAGG	745
MDR1HUMA.SEQ	AAAAATTTAGAAGAAGCTAAAAGAATTGGGATAAAGAAG	757
		
	CXATTXCXGCCAACATXGCCATAGGCTTCXCCTTCCTGAT	
	770 780 790 800	
mdr1FUGU.seq	GAGTTGCTGCTAACACGGCCACAGGCTTCTCCTTTCTGAT	776
mdr2FUGU.seq	CCATCTCCTCCAACATCGCCATGGGCTTCACCTTCCTGAT	785
MDR1HUMA.SEQ	CTATTACAGCCAATATTTCTATAGGTGCTGCTTTCCTGCT	797
		
	GATCTACCTGTCCTATGCTCTGGCCTTCTGGTACGGGACT	
	810 820 830 840	
mdr1FUGU.seq	GATCTACCTGTCCTATGCTCTGGCCTTCTGGTACGGGACT	816
mdr2FUGU.seq	GATCTACCTGTCCTATGCTCTGGCCTTCTGGTACGGGAGT	825
MDR1HUMA.SEQ	GATCTATGCATCTTATGCTCTGGCCTTCTGGTATGGGACC	837
		
	ACXCTGGTCCTCAAXXXXGAGTACACCATTGGAAATTTAC	
	850 860 870 880	
mdr1FUGU.seq	ACTCTGGTCCTCAACAAAGAGTACACCATTGGAAATTTAC	856
mdr2FUGU.seq	ACGCTCATCCTGAATTTTGGAGTACACCATCGGCAATTTAC	865
MDR1HUMA.SEQ	ACCTTGGTCCTCTCAGGGGAATATTTCTATTGGACAAGTAC	877
		
	TGACTGTGTTCTTTGTCGTXCTXATTGGGGCXTTTCAGTGT	
	890 900 910 920	
mdr1FUGU.seq	TGACTGTGTTCTTCGTCGTCCTCTACGGGGCATAACATTAT	896
mdr2FUGU.seq	TGACTGTGTTTTTTGTCGTCGTCCTTATTGGAGCGTTTCAGCGT	905
MDR1HUMA.SEQ	TCACTGTATTCTTTTCTGTATTAATTGGGGCTTTTAGTGT	917
		
	TGGACAGGCCTCTCCXAACATCCAGXCXTTTGCCAGTGCC	
	930 940 950 960	
mdr1FUGU.seq	TGGACAGGCCTCTCCCAACGTCCAGTCCTTTGCCAGTGCC	936
mdr2FUGU.seq	CGGACAGACCTCTCCGAACATCCAGAATTTTGCCAGCGCC	945
MDR1HUMA.SEQ	TGGACAGGCATCTCCAAGCATTGAAGCATTTGCAAATGCA	957

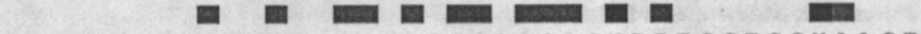
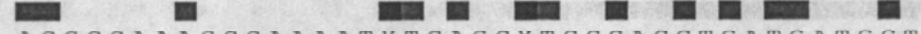
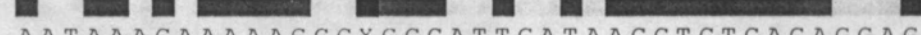
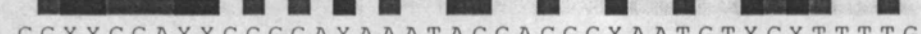
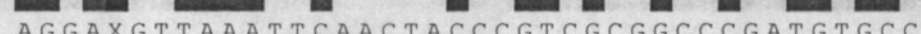
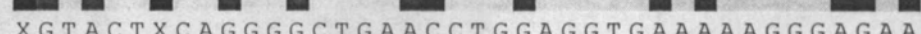
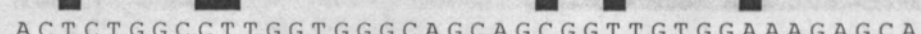
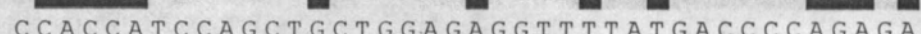
		
	CCTGAGAGAAATGATCGGAGTGGTCAGCCAGGAGCCCGTT	
	1290 1300 1310 1320	
mdr1FUGU.seq	CCTGAGAGAAATGATCGGAGTGGTCAGCCAGGAGCCCGTT	1296
mdr2FUGU.seq	CCTGAGAGAGATGATCGGAGTGGTCAGCCAGGAGCCCGTT	1305
MDR1HUMA.SEQ	TCTACGGGAAATCATTGGTGTGGTGAGTCAGGAACCTGTA	1317
		
	CTTTTCGCCACCACCATCACCGAGAACATCAGATACGGCC	
	1330 1340 1350 1360	
mdr1FUGU.seq	CTTTTCGCCACCACCATCACCGAGAACATCAGATACGGCC	1336
mdr2FUGU.seq	CTTTTCGCCACCACCATCACCGAGAACATCAGATACGGCC	1345
MDR1HUMA.SEQ	TTGTTTGCCACCACGATAGCTGAAAACATTCGCTATGGCC	1357
		
	GACTGGACGTGACGCAAGAGGAGATCGAACGAGCCACTAA	
	1370 1380 1390 1400	
mdr1FUGU.seq	GACTGGACGTGACGCAAGAGGAGATCGAACGAGCCACTAA	1376
mdr2FUGU.seq	GACTGGACGTGACGCAAGAGGAGATCGAACGAGCCACTAA	1385
MDR1HUMA.SEQ	GTGAAAATGTCACCATGGATGAGATTGAGAAAGCTGTCAA	1397
		
	AGAGTCCAACGCTTATGACTTCATCATGAACCTTCCAGAC	
	1410 1420 1430 1440	
mdr1FUGU.seq	AGAGTCCAACGCTTATGACTTCATCATGAACCTTCCAGAC	1416
mdr2FUGU.seq	AGAGTCCAACGCTTATGACTTCATCATGAACCTTCCAGAC	1425
MDR1HUMA.SEQ	GGAAGCCAATGCCTATGACTTTATCATGAAACTGCCTCAT	1437
		
	AAGTTTGAGACGCTGGTGGGAGATCGAGGGACTCAGCTGA	
	1450 1460 1470 1480	
mdr1FUGU.seq	AAGTTTGAGACGCTGGTGGGAGATCGAGGGACTCAGCTGA	1456
mdr2FUGU.seq	AAGTTTGAGACGCTGGTGGGAGATCGAGGGACTCAGCTGA	1465
MDR1HUMA.SEQ	AAATTTGACACCCTGGTTGGAGAGAGAGGGGCCAGTTGA	1477
		
	GCGGAGGACAGAAGCAGAGGATCGCCATCGCTCGAGCTCT	
	1490 1500 1510 1520	
mdr1FUGU.seq	GCGGAGGACAGAAGCAGAGGATCGCCATCGCTCGAGCTCT	1496
mdr2FUGU.seq	GCGGAGGACAGAAGCAGAGGATCGCCATCGCTCGAGCTCT	1505
MDR1HUMA.SEQ	GTGGTGGGCAGAAGCAGAGGATCGCCATTGCACGTGCCCT	1517
		
	GGTCCGCAACCCTAAAAATCCTCCTGCTGGACGAGGCCACG	
	1530 1540 1550 1560	
mdr1FUGU.seq	GGTCCGCAACCCTAAAAATCCTCCTGCTGGACGAGGCCACG	1536
mdr2FUGU.seq	GGTCCGCAACCCTAAAAATCCTCCTGCTGGACGAGGCCACG	1545
MDR1HUMA.SEQ	GGTTCGCAACCCCAAGATCCTCCTGCTGGATGAGGCCACG	1557
		
	TCTGCCCTCGATGCTGAGAGCGAGACCATCGTACAGGCTG	
	1570 1580 1590 1600	
mdr1FUGU.seq	TCTGCCCTCGATGCTGAGAGCGAGACCATCGTACAGGCTG	1576
mdr2FUGU.seq	TCTGCCCTCGATGCTGAGAGCGAGACCATCGTACAGGCTG	1585
MDR1HUMA.SEQ	TCAGCCTTGGACACAGAAAGCGAAGCAGTGGTTCAGGTGG	1597

		
	CTCTGGACAAGGTCCGACTGGGTTCGCACCACCATCGTGAT	
	1610 1620 1630 1640	
mdr1FUGU.seq	CTCTGGACAAGGTCCGACTGGGTTCGCACCACCATCGTCAT	1616
mdr2FUGU.seq	CTCTGGACAAGGTCCGACTGGGTTCGCACCACCATCGTGGT	1625
MDR1HUMA.SEQ	CTCTGGATAAAGGCCAGAAAAGGTTCGGACCACCATTTGTGAT	1637
		
	CGCTCACCGACTCTCGACCATCAGAAACGCCGACATCATT	
	1650 1660 1670 1680	
mdr1FUGU.seq	CGCTCACCGACTCTCGACCATCAGAAACGCCGACATCATT	1656
mdr2FUGU.seq	CGCTCACCGACTCTCGACCATCAGAAACGCCGACATCATT	1665
MDR1HUMA.SEQ	AGCTCATCGTTTGTCTACAGTTTCGTAATGCTGACGTCATC	1677
		
	GCTGGATTTCAGTAATGGXGAAATCGTGGAGCAGGGGACTC	
	1690 1700 1710 1720	
mdr1FUGU.seq	GCTGGATTTCAGTAATGGTGAATCGTTGAGCAGGGGACTC	1696
mdr2FUGU.seq	GCTGGATTTCAGTAATGGCAAAATCGTGGAGCAGGGGACTC	1705
MDR1HUMA.SEQ	GCTGGTTTTCGATGATGGAGTCATTTGTGGAGAAAGGAAATC	1717
		
	ACAGCCAGCTGATGGAGATAAAGGGAGTCTATCATGGCCT	
	1730 1740 1750 1760	
mdr1FUGU.seq	ACAGCCAGCTGATGGAGATAAAGGGAGTCTATCATGGCCT	1736
mdr2FUGU.seq	ACAGCCAGCTGATGGAGATAAAGGGAGTCTATCATGGCCT	1745
MDR1HUMA.SEQ	ATGATGAACTCATGAAAGAGAAAGGCATTTACTTCAAAC	1757
		
	GGTGACCATGCAGACXTTXCAXAATGTGGAGGAXXTAGAA	
	1770 1780 1790 1800	
mdr1FUGU.seq	GGTGACCATGCAGAGCTTTCAGAAGCTGGAGGATCTGGAA	1776
mdr2FUGU.seq	GGTGACCATGCAGACGTTCCACAATGTGGAGGAGGAAAAT	1785
MDR1HUMA.SEQ	TGTCACAATGCAGACAGCAGGAAATGAAGTTGAATTAGAA	1797
		
	AACGCAGXXTAXGAGTCCTXXGXTGXGAGAAGAGCCXGG	
	1810 1820 1830 1840	
mdr1FUGU.seq	GA CTCAGACTACGAGCCCTGGGTTCGCTGAGAAGAGCCAGC	1816
mdr2FUGU.seq	ACCGCCATGTCGGAGTTATCTGCTGGGGAGAAGAGCCCTG	1825
MDR1HUMA.SEQ	AATGCAGCTGATGAATCCAAAAGTGAAATTGATGCCTTGG	1837
		
	TGATXXAXTC --- CGTCTCCAGTCCTCCCTXATXAGGAG	
	1850 1860 1870 1880	
mdr1FUGU.seq	TGATCGAATC --- CTTCTCCAGTCCTCCCTGCAGAGGAG	1853
mdr2FUGU.seq	TGGAAAAGAC --- CGTCTCCAGTCGTCCATCATCAGGAG	1862
MDR1HUMA.SEQ	AAATGTCTTCAAATGATTCAAGATCCAGTCTAATAAGAAA	1877
		
	GAGGTCCACTAGAGGGTCTTGGCTTGCTGCTCAGAAGGA	
	1890 1900 1910 1920	
mdr1FUGU.seq	GAGGTCCACTAGAGGCTCTTGGCTTGCTGCTCAGAAGGA	1893
mdr2FUGU.seq	GAAGTCCACCAGAGGGTCTCGTTTGCCGCGTCAGAAGGA	1902
MDR1HUMA.SEQ	AAGATCAACTCGTAGGAGTGTCCGTGGATCACAAAGCCAA	1917

	ACCAAAGAGGAXAAX-----	
	1930 1940 1950 1960	
mdr1FUGU.seq	ACAAAAGAGGAGAAAGAAAATTTGAGTGCGATCAGGACA	1933
mdr2FUGU.seq	ACCAAAGAGGAAAAG-----	1918
MDR1HUMA.SEQ	GACAGAAAGCTTAGT-----	1933
	--ACAGAAGAGG-----ATGAAAACGTTCCCTCCCGTGTC	
	1970 1980 1990 2000	
mdr1FUGU.seq	ACATAGAGGAGG-----ATGAGAACGTTCCCTCCCGTGTC	1967
mdr2FUGU.seq	--ACAGAAGAGG-----ATGAAGACGTTCCCGACGTGTC	1949
MDR1HUMA.SEQ	--ACCAAAGAGGCTCTGGATGAAAGTATACCTCCAGTTTC	1970
	GTTCTTTAAAGTGATGCATCTXAACXTXXCTGAGTGGCCX	
	2010 2020 2030 2040	
mdr1FUGU.seq	GTTCTTTAAAGTGATGCGTTACAACGTTTCTGAGTGGCCG	2007
mdr2FUGU.seq	GTTCTTTAAAGTGCTGCATCTGAACATCCCCGAGTGGCCC	1989
MDR1HUMA.SEQ	CTTTTGGAGGATTATGAAGCTAAATTTAACTGAATGGCCT	2010
	TATATTXTTGTXGGXXTCATCTGCGCCATGATCAATGGAG	
	2050 2060 2070 2080	
mdr1FUGU.seq	TATATTTTGGTAGGAACCATCTGCGCCATGATCAACGGTG	2047
mdr2FUGU.seq	TACATCCTTGTGGGGCTCATCTGCGCTACGATCAATGGAG	2029
MDR1HUMA.SEQ	TATTTTGTGTGTGGTGTATTTTGTGCCATTATAAATGGAG	2050
	CCATGCAGCCAGTXTTCGCCATCATCTTCTCCAAGATCAT	
	2090 2100 2110 2120	
mdr1FUGU.seq	CGATGCAGCCAGTGTTTCAGCATCATCTTCACCGAGATCAT	2087
mdr2FUGU.seq	CCATGCAGCCGGTCTTCGCCATCCTCTTCTCCAAGATCAT	2069
MDR1HUMA.SEQ	GCCTGCAACCAGCATTGTGCAATAATATTTTCAAAGATTAT	2090
	XAXGGTXTTTACAGAAAXXGATXAT---GAAACTATCCGG	
	2130 2140 2150 2160	
mdr1FUGU.seq	TATGGTCTTCAGAGAAAAGGATAAA---GAAATTATCCGG	2124
mdr2FUGU.seq	CACTGTGTTTGCGGATCCAGACCGT---GATTCTGTTCAGG	2106
MDR1HUMA.SEQ	AGGGGTTTTTTACAAGAATTGATGATCCTGAAACAAAACGA	2130
	XAGAAGAGTXAATTCTTTTCTCTAXTGTGTTTGTGTXGTXATTG	
	2170 2180 2190 2200	
mdr1FUGU.seq	GAGAAGTCGTCATTCTTTTGTATACTGTTTGTGCTGTGATGG	2164
mdr2FUGU.seq	AGGAAGAGTGAATTCATTTCTCTGATGTTTGTGCTTATTG	2146
MDR1HUMA.SEQ	CAGAATAGTAACTTGTTTTCACTATTGTTTCTAGCCCTTG	2170
	GXXTTGTGTCTTTTXXTCACCATGTTT-----	
	2210 2220 2230 2240	
mdr1FUGU.seq	GGGTGGTGACTTTTCTCACCATGTTTCTGCAGTTTTGGGG	2204
mdr2FUGU.seq	GCTGTGTGTCATTTGTACCATGTTT-----	2173
MDR1HUMA.SEQ	GAATTATTTCTTTTATTACATTTTTC-----	2197

		
	- TTXCAGGGTTTCTGTTTCGGTAAATCTGGAGAGATTCTG	
	2250 2260 2270 2280	
mdr1FUGU.seq	TTTCCAGGGTTTCTGTTTCAGTAAATCTGGAGAAATTCTG	2244
mdr2FUGU.seq	- TTACAGGGTTACTGTTTCGGTAAATCCGGAGAGATTCTG	2211
MDR1HUMA.SEQ	- CTCAGGGTTTTCACATTTGGCAAAGCTGGAGAGATCCTC	2235
		
	ACCCTGAAGCTCAGACTCAXGGCXTTCAXATCXATGATGA	
	2290 2300 2310 2320	
mdr1FUGU.seq	ACCCTGAACCTCAGACTCAAAGCCTTCATATCTATGATGA	2284
mdr2FUGU.seq	ACGCTGAAGCTGAGACTCCGGGCGTTCACGGCGATGATGA	2251
MDR1HUMA.SEQ	ACCAAGCGGCTCCGATACATGGTTTTCCGATCCATGCTCA	2275
		
	GACAGGACCTCAGCTGGTACGACAATCCCAAAAACACCGT	
	2330 2340 2350 2360	
mdr1FUGU.seq	GACAGGACCTCAGCTGGTACGACAATCCCAAAAACACCGT	2324
mdr2FUGU.seq	GACAGGACCTCAGCTGGTACGACAATCCCAAAAACACCGT	2291
MDR1HUMA.SEQ	GACAGGATGTGAGTTGGTTTTGATGACCCTAAAAACACCAC	2315
		
	TGGCGCTCTCACCCTAGGCTGGCTGCCGACGCCGCCCAA	
	2370 2380 2390 2400	
mdr1FUGU.seq	TGGCGCTCTCACCCTAGGCTGGCTGCCGACGCCGCCCAC	2364
mdr2FUGU.seq	TGGCGCTCTCACCCTAGGCTGGCTGCCGACGCCGCCCAA	2331
MDR1HUMA.SEQ	TGGAGCATTGACTACCAGGCTCGCCAATGATGCTGCTCAA	2355
		
	GTACAAGGAGCTGCAGGGGTGCGCCTGGCTGTAATXACGC	
	2410 2420 2430 2440	
mdr1FUGU.seq	GTACAAGGAGCTGCAGGGGTGCGCCTGGCTGTAATGACGC	2404
mdr2FUGU.seq	GTACAAGGAGCTGCAGGGGTGCGCCTGGCGACAATAATGC	2371
MDR1HUMA.SEQ	GTTAAAGGGGCTATAGGTTCCAGGCTTGCTGTAATTACCC	2395
		
	AGAACTTCGCCAACCTGGGCACCAGCATCATCATCXCTT	
	2450 2460 2470 2480	
mdr1FUGU.seq	AGAACTTCGCCAACCTGGGCACCAGCATCATCATCAGCTT	2444
mdr2FUGU.seq	AGAACTTCGCCAACCTGGGCACCAGCATCATCATCGCTT	2411
MDR1HUMA.SEQ	AGAATATAGCAAATCTTGGGACAGGAATAATTATATCCTT	2435
		
	CGTXTACGGCTGGGAGCTGACCCTGCTCATCCTGGCCGTG	
	2490 2500 2510 2520	
mdr1FUGU.seq	CGTGTACGGCTGGGAGCTGACCCTGCTCATCCTGGCCGTG	2484
mdr2FUGU.seq	TGTTTACGGCTGGGAGCTGACCCTGCTCATCCTGGCCGTG	2451
MDR1HUMA.SEQ	CATCTATGGTTGGCAACTAACACTGTTACTCTTAGCAATT	2475
		
	GTGCCCATCATXGXCXGTXGCCGGAGCCGCTGAGATCAAGC	
	2530 2540 2550 2560	
mdr1FUGU.seq	GTGCCCATCCTGGCTGTGGCCGGAGCCGCTGAGGTCAAGC	2524
mdr2FUGU.seq	GTGCCCTCATCGCGGCCGCCGGAGCCGCTGAGATCAAGC	2491
MDR1HUMA.SEQ	GTACCCATCATTTGCAATAGCAGGAGTTGTTGAAATGAAA	2515

		TGCTGXCXGGACACGCCGCAAAGACAAGAAGGAGCTGGA	
		2570 2580 2590 2600	
mdr1FUGU.seq		TGCTGACAGGACACGCCGCAAAGACAAGAAGGAGCTGGA	2564
mdr2FUGU.seq		TGCTCGCGGGTACGCCGCAAAGACAAGAAGGAGCTGGA	2531
MDR1HUMA.SEQ		TGTTGTCTGGACAAGCACTGAAAGATAAGAAAGAAGCTAGA	2555
		GAXGGCCGGAAAGATCGCCACAGAGGCCATCGAGAACGTC	
		2610 2620 2630 2640	
mdr1FUGU.seq		GATGGCCGGAAAGATCGCCACAGAGGCCATCGAGAATGTG	2604
mdr2FUGU.seq		GAAGGCCGGAAAGATCGCCACAGAGGCCATCGAGAACGTC	2571
MDR1HUMA.SEQ		AGGTGCTGGGAAGATCGCTACTGAAGCAATAGAAAAGTTTTC	2595
		AGAACCGTXGTGTCCCTCACCAGAGAACCGAAATTTGAGX	
		2650 2660 2670 2680	
mdr1FUGU.seq		AGAACTGTGGTGTCCCTCACCAGAGAACCGACATTTGTGG	2644
mdr2FUGU.seq		AGAACCGTCGTGTCCCTCAGCAGAGAACCAAAATTTGAGT	2611
MDR1HUMA.SEQ		CGAACCGTTGTTTTCTTTGACTCAGGAGCAGAAGTTTGAAC	2635
		XTTTATATGAGGAGAATCTXAXXGTXCCATACAAGAACTC	
		2690 2700 2710 2720	
mdr1FUGU.seq		CTTTATACGAGGAAAATCTAACTGTTCCATACAAGAACTC	2684
mdr2FUGU.seq		GTTTATATGAGGAGAATCTCAGAGTGCCGTACAAGAACTC	2651
MDR1HUMA.SEQ		ATATGTATGCTCAGAGTTTGCAGGTACCATACAGAAACTC	2675
		CCAGAAAAAGGCCACATXTATGGCTTAACCTACTCCTTC	
		2730 2740 2750 2760	
mdr1FUGU.seq		CCAGAAAAAGGCCAAAATTTATGGCTTAACCTACTCCTTC	2724
mdr2FUGU.seq		CCAGAAAAAGGCCACGTGTACGGCTTAACCTACTCCTTC	2691
MDR1HUMA.SEQ		TTTGAGGAAAGCACACATCTTTGGAATTACATTTTTCCTTC	2715
		TCCCAGGCCATGATCTACTTTGCTTACGCTGCCTGTTTCC	
		2770 2780 2790 2800	
mdr1FUGU.seq		TCACAGGCCATGATCTTCTTTGTTTACGCTGCCTGTTTCC	2764
mdr2FUGU.seq		TCCCAGGCCATGATCTACTTTGCTTACGCTGCCTGTTTCC	2731
MDR1HUMA.SEQ		ACCCAGGCAATGATGTATTTTTCCTATGCTGGATGTTTCC	2755
		GCTTTGGAGCCTGGCTGATXGAAGCAGGACGGATGGACGT	
		2810 2820 2830 2840	
mdr1FUGU.seq		GCTTTGGAGCCTGGCTGATCGAAGCAGGACGGATGGATGT	2804
mdr2FUGU.seq		GCTTCGGAGCCTGGCTGATTGAAGCAGGACGGATGGACGT	2771
MDR1HUMA.SEQ		GGTTTGGAGCCTACTTGGTGGCACATAAACTCATGAGCTT	2795
		GGAGGGAGTGTTCCCTXGTGGTTTTCXGCXGTGCTGTACGGC	
		2850 2860 2870 2880	
mdr1FUGU.seq		GGAGGGAGTGTTCCCTTGTGGTTATGACAATGCTGTACGGC	2844
mdr2FUGU.seq		GGAGGGAGTGTTCCCTGGTGGTTTCTGCGGTGCTGTACGGC	2811
MDR1HUMA.SEQ		TGAGGATGTTCTGTTAGTATTTTTCAGCTGTTGTCTTTGGT	2835

		
	GCCATGGCCGTGGGGGAAGCCAACACXTTGTGCTCCXAACT	
	2890 2900 2910 2920	
mdr1FUGU.seq	GCCATGGCTGTGCGGCGAGGCCAACACTTATGCTCCCAACT	2884
mdr2FUGU.seq	GCCATGGCCGTGGGGGAAGCTAACACCTTGTGCTCCGAACT	2851
MDR1HUMA.SEQ	GCCATGGCCGTGGGGCAAGTCAGTTCATTTGCTCCTGACT	2875
		
	ACGCCAAAGCCAAAATXTCAGCXTCCACCTGATGATGCT	
	2930 2940 2950 2960	
mdr1FUGU.seq	TCGCCAAAGCCAAAATCTCAGCCTCCACCTGACGATGCT	2924
mdr2FUGU.seq	ACGCCAAGGCCAAAATGGCTGCTTCCTACCTGATGATGCT	2891
MDR1HUMA.SEQ	ATGCCAAAGCCAAAATATCAGCAGCCACATCATCATGAT	2915
		
	AATAAACAAAAAGCCXGCCATTGATAACCTCTCAGAGGAG	
	2970 2980 2990 3000	
mdr1FUGU.seq	AATAAACAGACAGCCGGCCATAGATAATCTGTGTCAGAGGAG	2964
mdr2FUGU.seq	AATAACAAGAAGCCCGCCATTGATAACCTCTCAGAGGAG	2931
MDR1HUMA.SEQ	CATTGAAAAACCCCTTTGATTGACAGCTACAGCACGGAA	2955
		
	GGXXCGAXXCCGGAXAAATACGACGGXAATGTXCXTTTTG	
	3010 3020 3030 3040	
mdr1FUGU.seq	GAAGCGAGACTGGAGAAATACGACGGCAACGTTCTTTTGT	3004
mdr2FUGU.seq	GGGACGTCTCCGGAAAAATACGACGGTAATGTGCATTTCTG	2971
MDR1HUMA.SEQ	GGCCTAATGCCGAACACATTGGAAGGAAATGTCACATTTG	2995
		
	AGGAXGTTAAATTCAACTACCCGTCGCGGCCCGATGTGCC	
	3050 3060 3070 3080	
mdr1FUGU.seq	AGGACGTCAAGTTTAACTACCCGTCGCGGCCCGATGTGCC	3044
mdr2FUGU.seq	AGGGTGTAAATTCAACTACCCGTCGCGGCCCGATGTGAC	3011
MDR1HUMA.SEQ	GTGAAGTTGTATTCAACTATCCACCCGACCGGACATCCC	3035
		
	XGTACTXCAGGGGCTGAACCTGGAGGTGAAAAAGGGAGAA	
	3090 3100 3110 3120	
mdr1FUGU.seq	TGTACTACAAGGGCTGAATCTGGAGGTGCAAAGGGAGAA	3084
mdr2FUGU.seq	CATACTCCAGGGGCTGAACCTGAAGGTGAAAAAGGGAGAA	3051
MDR1HUMA.SEQ	AGTGCTTCAGGGACTGAGCCTGGAGGTGAAGAAGGGCCAG	3075
		
	ACTCTGGCCTTGGTGGGCAGCAGCGGTTGTGGAAAGAGCA	
	3130 3140 3150 3160	
mdr1FUGU.seq	ACTCTGGCCTTGGTGGGCAGCAGCGGTTGTGGAAAGAGCA	3124
mdr2FUGU.seq	ACTCTGGCCTTGGTGGGCAGCAGCGGTTGTGGAAAGAGCA	3091
MDR1HUMA.SEQ	ACGCTGGCTCTGGTGGGCAGCAGTGGCTGTGGGAAGAGCA	3115
		
	CCACCATCCAGCTGCTGGAGAGGTTTTATGACCCAGAGA	
	3170 3180 3190 3200	
mdr1FUGU.seq	CCACCATCCAGCTGCTGGAGAGGTTTTATGACCCAGAGA	3164
mdr2FUGU.seq	CCACCATCCAGCTGCTGGAGAGGTTTTATGACCCAGAGA	3131
MDR1HUMA.SEQ	CAGTGGTCCAGCTCCTGGAGCGGTTCTACGACCCCTTGGC	3155

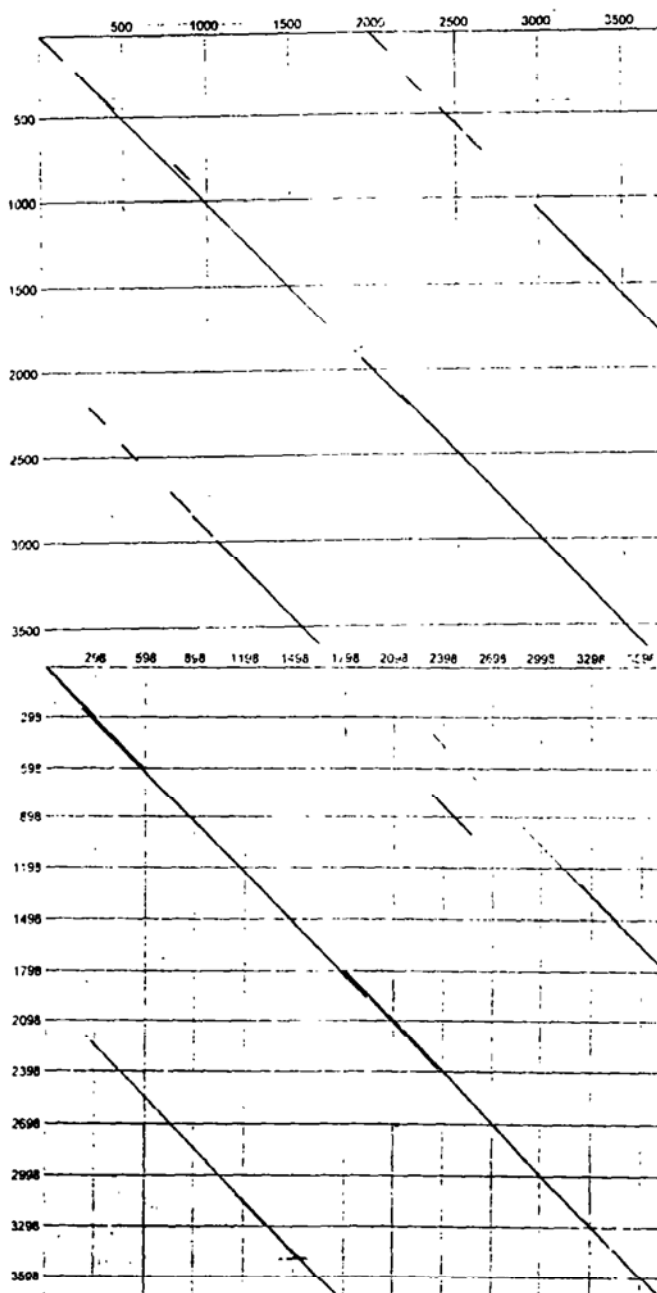
	GGGGAGAGTGTGCTGGACGGTGTGCGAXGTGAAACAGCTG	
	3210 3220 3230 3240	
mdr1FUGU.seq	GGGGAGAGTGTGCTGGACGGTGTGCGATGTGAAACAGCTG	3204
mdr2FUGU.seq	GGGGAGAGTGTCACTGGACGGTGTCAACGTGAAACAGCTG	3171
MDR1HUMA.SEQ	AGGGAAAGTGCTGCTTGATGGCAAAGAAATAAAGCGACTG	3195
	AACGTTCACTGGCTGAGGTCTCAGATCGGCATCGTCTCCC	
	3250 3260 3270 3280	
mdr1FUGU.seq	AACGTTCACTGGCTGAGGTCTCAGATCGGCATCGTCTCCC	3244
mdr2FUGU.seq	AACATTTCACTGGCTGAGGTCTCAGATCGGCATCGTCTCCC	3211
MDR1HUMA.SEQ	AATGTTCACTGGCTCCGAGCACACCTGGGCATCGTGTCCC	3235
	AGGAGCCGGTGCTGTTGCTGACTGCTCCCTGGCTGAGAACAT	
	3290 3300 3310 3320	
mdr1FUGU.seq	AGGAGCCGGTGCTGTTGCTGACTGCTCCCTGGCTGAGAACAT	3284
mdr2FUGU.seq	AGGAGCCGGTGCTGTTGCTGACTGCTCCCTGGCTGAGAACAT	3251
MDR1HUMA.SEQ	AGGAGCCCATCCTGTTTACTGCAGCATTTGCTGAGAACAT	3275
	CGCCTACGGAGACAACAGTTCGCTCCGTGTCCATGGATGAG	
	3330 3340 3350 3360	
mdr1FUGU.seq	CGCCTACGGAGACAACAGTTCGCTCCGTGTCCATGGATGAG	3324
mdr2FUGU.seq	CGCCTACGGAGACAACAGTTCGCTCCGTGTCCATGGATGAG	3291
MDR1HUMA.SEQ	TGCCTATGGAGACAACAGCCGGGTGGTGTACAGGAAGAG	3315
	ATAGTAXXTGCTGCTAAAGCAGCCAACATCCACAGCTTCA	
	3370 3380 3390 3400	
mdr1FUGU.seq	ATAGTAGCTGCTGCTAAAGCAGCCAACATCCACAGCTTCA	3364
mdr2FUGU.seq	ATAGTACTTGCTGCTAAAGCAGCCAACATCCACAGCTTCA	3331
MDR1HUMA.SEQ	ATCGTGAGGGCAGCAAAGGAGGCCAACATACATGCCTTCA	3355
	TCGAAGGGCTGCCTCAGAGATACGACACTCAGGCTGGTGA	
	3410 3420 3430 3440	
mdr1FUGU.seq	TCGAAGGGCTGCCTCAGAGATACGACACTCAGGCTGGTGA	3404
mdr2FUGU.seq	TCGAAGGGCTGCCTCAGAGATACGACACTCAGGCTGGTGA	3371
MDR1HUMA.SEQ	TCGAGTCACTGCCTAATAAATATAGCACTAAAGTAGGAGA	3395
	TAAGGGAACACAGCTGTCAGGGGGCCAGAAGCAGCGTGTC	
	3450 3460 3470 3480	
mdr1FUGU.seq	TAAGGGAACACAGCTGTCAGGGGGCCAGAAGCAGCGTGTC	3444
mdr2FUGU.seq	TAAGGGAACACAGCTGTCAGGGGGCCAGAAGCAGCGTGTC	3411
MDR1HUMA.SEQ	CAAAGGAACACTCAGCTCTCTGGTGGCCAGAACAACGCATT	3435
	GCCATAGCCCAGCCATCATCCGCAACCCCAAACCTGTTGC	
	3490 3500 3510 3520	
mdr1FUGU.seq	GCCATAGCCCAGCCATCATCCGCAACCCCAAACCTGTTGC	3484
mdr2FUGU.seq	GCCATAGCCCAGCCATCATCCGCAACCCCAAACCTGTTGC	3451
MDR1HUMA.SEQ	GCCATAGCTCGTGCCCTTGTAGACAGCCTCATATTTTGC	3475

	TCCTGGACGAGGCCACGTCTGCGCTCGACACTGAGAGTGA	
	3530 3540 3550 3560	
mdr1FUGU.seq	TCCTGGACGAGGCCACGTCTGCGCTCGACACTGAGAGTGA	3524
mdr2FUGU.seq	TCCTGGACGAGGCCACGTCTGCGCTCGACACTGAGAGTGA	3491
MDR1HUMA.SEQ	TTTTGGATGAAGCCACGTCTGCTCTGGATAACAGAAAGTGA	3515
	GAAGGTGGTGCAGGAGGCCGTTGGACCAGGCCAGGAAGGGC	
	3570 3580 3590 3600	
mdr1FUGU.seq	GAAGGTGGTGCAGGAGGCCGTTGGACCAGGCCAGGAAGGGC	3564
mdr2FUGU.seq	GAAGGTGGTGCAGGAGGCCGTTGGACCAGGCCAGGAAGGGC	3531
MDR1HUMA.SEQ	AAAGGTTGTCCAAGAAGCCCTGGACAAAGCCAGAGAAGGC	3555
	AGGACGTGCATCGTGGTAGCCACCGTCTGTCCACCATCC	
	3610 3620 3630 3640	
mdr1FUGU.seq	AGGACGTGCATCGTGGTAGCCACCGTCTGTCCACCATCC	3604
mdr2FUGU.seq	AGGACGTGCATCGTGGTAGCCACCGTCTGTCCACCATCC	3571
MDR1HUMA.SEQ	CGCACCTGCATTGTGATTGCTCACCGCCTGTCCACCATCC	3595
	AGAACGCCGACTGCATCGCTGTGTTCCAGGGAGGAGTGGT	
	3650 3660 3670 3680	
mdr1FUGU.seq	AGAACGCCGACTGCATCGCTGTGTTCCAGGGAGGAGTGGT	3644
mdr2FUGU.seq	AGAACGCCGACTGCATCGCTGTGTTCCAGGGAGGAGTGGT	3611
MDR1HUMA.SEQ	AGAATGCAGACTTAATAGTGGTGTTCAGAATGGCAGAGT	3635
	GGTGGAAAAGGGGACGCACCAGCAGCTGATCGCCAAGAAG	
	3690 3700 3710 3720	
mdr1FUGU.seq	GGTGGAAAAGGGGACGCACCAGCAGCTGATCGCCAAGAAG	3684
mdr2FUGU.seq	GGTGGAAAAGGGGACGCACCAGCAGCTGATCGCCAAGAAG	3651
MDR1HUMA.SEQ	CAAGGAGCATGGCAGCATCAGCAGCTGCTGGCACAGAAA	3675
	GGAGTGTACCACATGCTGGTACCAAACAGATGGGCTATC	
	3730 3740 3750 3760	
mdr1FUGU.seq	GGAGTGTACCACATGCTGGTACCAAACAGATGGGCTATC	3724
mdr2FUGU.seq	GGAGTGTACCACATGCTGGTACCAAACAGATGGGCTATC	3691
MDR1HUMA.SEQ	GGCATCTATTTTTCAATGGTCAAGTGTCCAGGCTGGAACAA	3715
	<u>ACAGCGAX</u>	
mdr1FUGU.seq	ACAGTGGA	3732
mdr2FUGU.seq	ACAACGAC	3699
MDR1HUMA.SEQ	AGCGCCAG	3723

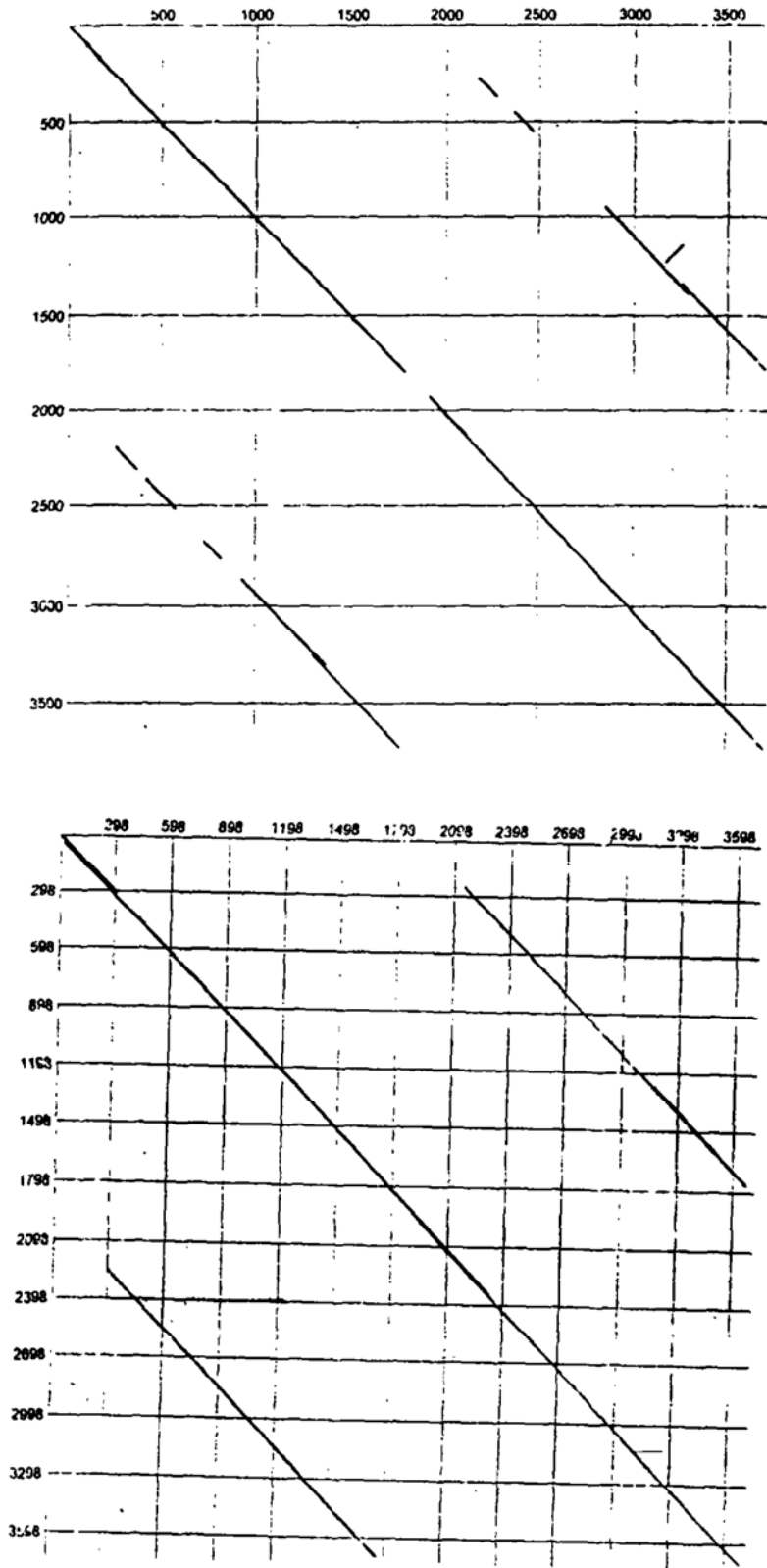
Percent Similarity					
	1	2	3		
1	■	84.7	61.4	1	mdr1FUGU.seq
2	17.2	■	61.9	2	mdr2FUGU.seq
3	55.0	54.1	■	3	mdr1HUMA.seq
	1	2	3		

dotplot 分析直观的表明三者之间有着极高的同源性，尤其是核苷酸结合位点。

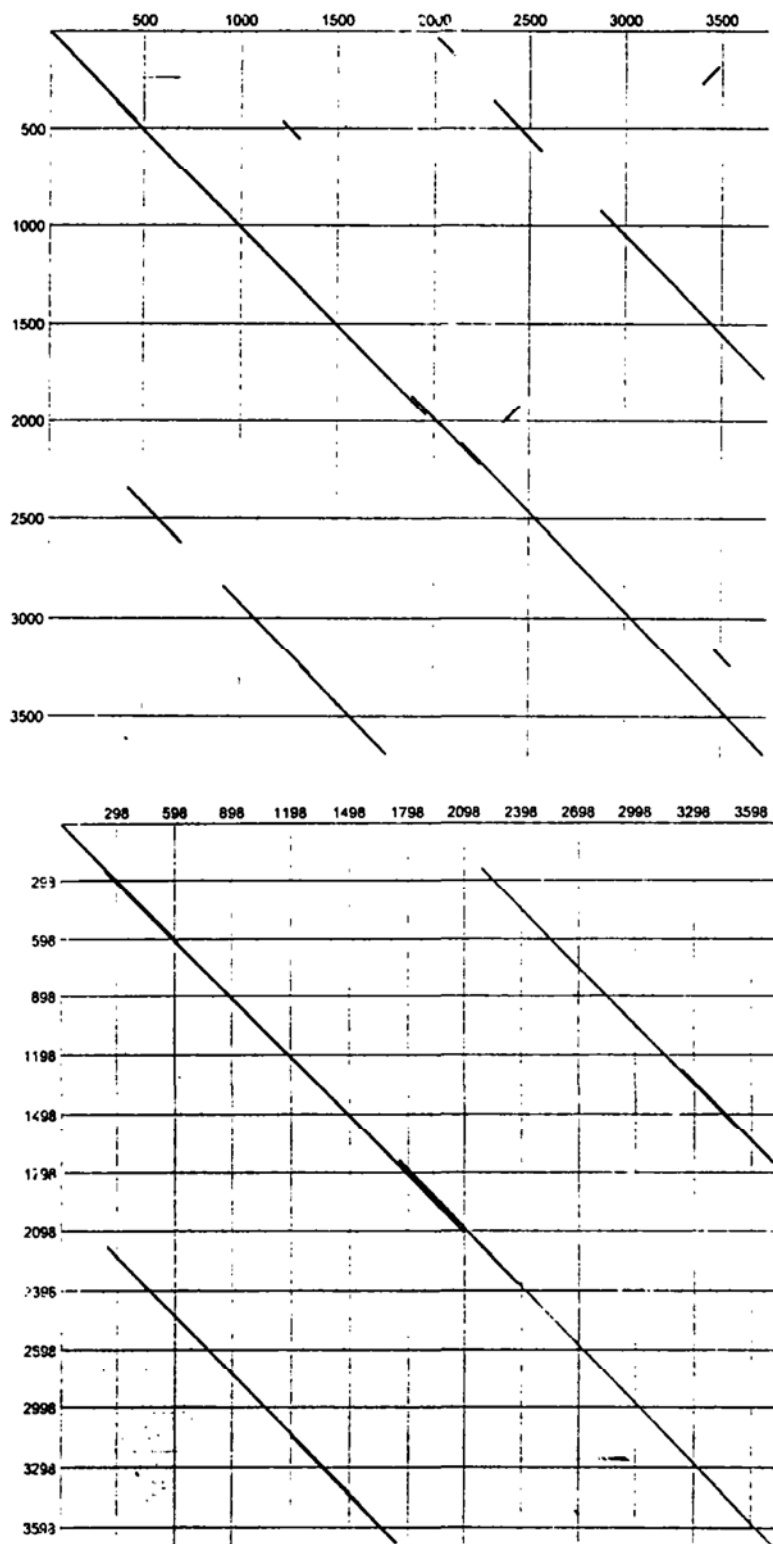
FuguMDR1 与 humanMDR1 核苷酸 (上) 和氨基酸 (下) 水平的 dotplot 结果



FuguMDR2 与 humanMDR1 核苷酸 (上) 和氨基酸 (下) 水平的 dotplot 结果



FuguMDR1 与 fuguMDR2 核苷酸 (上) 和氨基酸 (下) 水平的 dotplot 结果

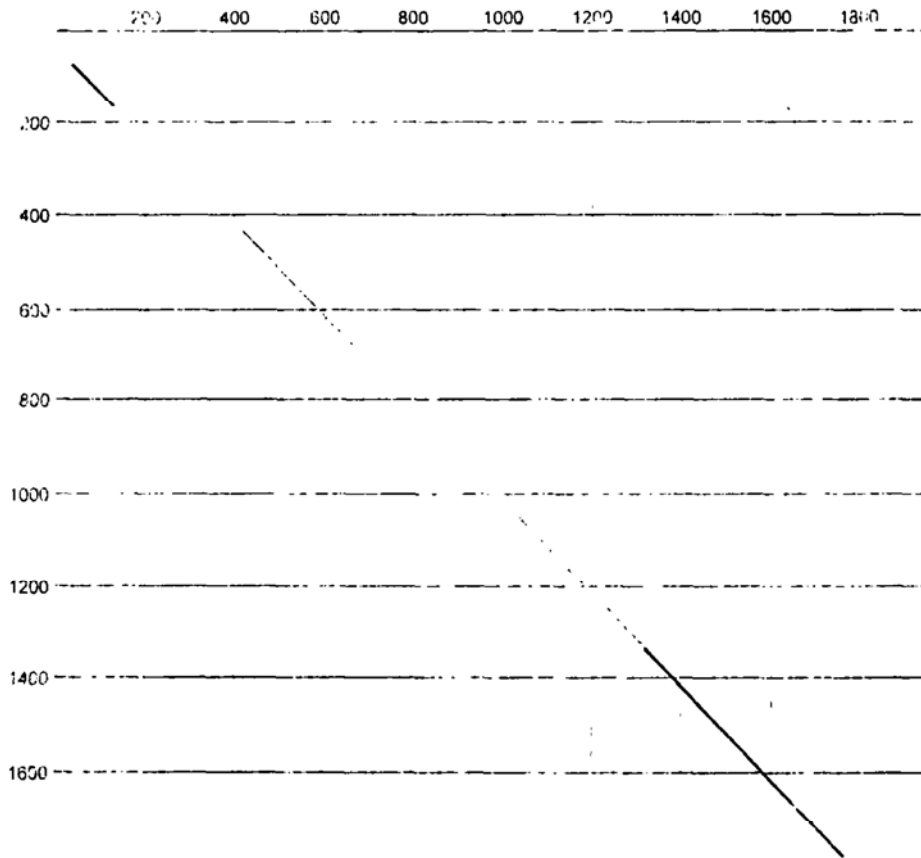


二 利用 Dnastar 软件对 fuguMDR1 cDNA 和 fuguMDR2 cDNA 的 5 端半侧与 3 端半侧和 N 端半侧与 C 端半侧的同源性进行比较

结果显示: 5 端半侧与 3 端半侧和 N 端半侧与 C 端半侧的同源性都较高, 尤其是核苷酸结合位点。

fuguMDR1 cDNA 的 5'端半侧和 3'端半侧的同源性比较结果:

Dotplot 结果: 横坐标为 5 端半侧; 纵坐标为 3 端半侧



TTTGGGTTTGTGCTGGTGTGTCGTTCTTTATGGTGATGC
 10 20 30 40

5mdr1FUG.seq TTCCGGTTTGCAGATGGTTGGGACATCTTGATGGT----- 36
 3mdr1FUG.seq GATGAGAACGTTTCTCCCGTGTGTCGTTCTTTAAAGTGATGC 40

GTTACACCGTTTGTGTGTTGCTGTCTATTTTGGTCGGGCT
 50 60 70 80

5mdr1FUG.seq -----CACCATCGGAGTGCTGATGGCCATTGTGAACGGGCT 71
 3mdr1FUG.seq GTTACAACGTTTCTGAGTGGCCGTATATTTTGGTAGGAAC 80

GGTGTGTGCTCTGATGTGTGTTGTGTTTGGTGCGGTGTTT
 90 100 110 120

5mdr1FUG.seq GGTGAATCCTCTGATGTGTATCGTGTTTGGTGAGATGACT 111
 3mdr1FUG.seq CATCTGCGCCATGATCAACGGTGCGATGCAGCCAGTGTTTC 120

GGCATCTTCTTCCCGGAGGTCATTATTGTCTTCAGAGCAC
 130 140 150 160

5mdr1FUG.seq GACAGCTTCATCCAGGAAGCCAA-ACTGTCCCAAACCAC 150
 3mdr1FUG.seq AGCATCATCTTCACCGAGATCATTATGGTCTTCAGAGAAA 160

AGGATAAGGAACTTCTCTTGGAGAATTTGTCATTTTTTTG
 170 180 190 200

5mdr1FUG.seq AACACAAGCAACCCCACTTCAACAATGTCTAATGTTTTAA 190
 3mdr1FUG.seq AGGATAAAGAAATTATCCGGGAGAAGTCGTCATTCTTTTG 200

GAGTGTGCTGTATTGTTTGCTGTGTTTTGTGTGTTTACTT
 210 220 230 240

5mdr1FUG.seq GAGTGTGCTGCATCATTGCGCCATCTATTCTCCATCTACTA 230
 3mdr1FUG.seq -----TATACTGTTTGCTGTGATGGGGGTGGTGACTT 232

TTCTCTCTTTGTTTTTGTGTTTTGGTGTTTGCGTGGTTT
 250 260 270 280

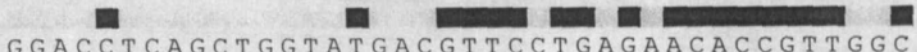
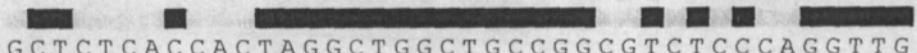

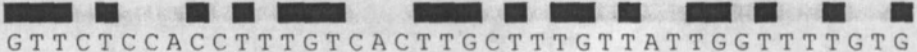


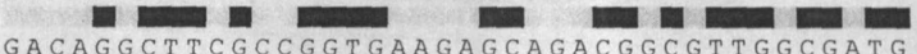
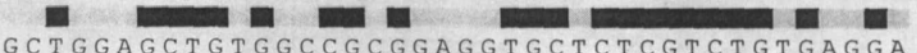
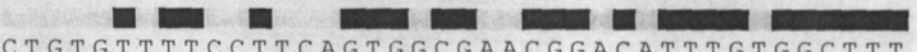
5mdr1FUG.seq CTCCATCTTGGGGTTTGTGCTGTGCTGGTAGTGGCGTACCTG 270
 3mdr1FUG.seq TTCTCACCATGTTTCTGCAGTTTTTGGGGTTTCCAGGGTTT 272

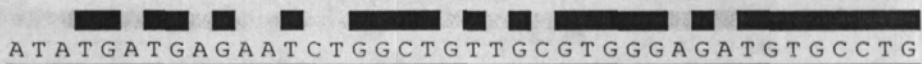
CAGATGTTTCTGTGGATCTTGAGGGCTTCTGGGCCTGCC
 290 300 310 320

5mdr1FUG.seq CAGATGTCTCTGTGGACCCTAACGGCCGCGCGGCAGGCCA 310
 3mdr1FUG.seq C---TGTTTCAGTAAATCTGGAGAAATTCTGACCCTGAAC 309

CTCGGATTTCGCGGGTTTTTTTTCTATGGCATCATGCGGCA
 330 340 350 360

5mdr1FUG.seq AACGAATTTCGCGAGTTGTTTTTCCACGGCATCATGCAGCA 350
 3mdr1FUG.seq CTCAGACTCAAAGCCTTCATATCTATG-----ATGAGACA 344

		
	GGACCTCAGCTGGTATGACGTTCTGAGAACACCGTTGGC	
	370 380 390 400	
5mdr1FUG.seq	GGACATCAGCTGGTATGACGTGACTGAGA-----CAGGA	384
3mdr1FUG.seq	GGACCTCAGCTGGTACGACAATCCCAAAAACACCGTTGGC	384
		
	GCTCTCACCACTAGGCTGGCTGCCGGCGTCTCCAGGTTG	
	410 420 430 440	
5mdr1FUG.seq	GAGCTCAACAC-----GCGTCTCACAGAGTG	410
3mdr1FUG.seq	GCTCTCACCACTAGGCTGGCTGCCGACGCCGCCACGTAC	424
		
	GGGGGGCTTCGGTGGTGCGGCTGGTTGTGCTGATGCAGGC	
	450 460 470 480	
5mdr1FUG.seq	GGAGGGCATCGGTGACAAGGCGGGTC-TGCTGATCCAGGC	449
3mdr1FUG.seq	AAGGAGCTGCAGGGGTGCGCCTGGCTGTAATGACGCAGAA	464
		
	GTTCTCCACCTTTGTCACTTGTCTTTGTTATTGGTTTTGTG	
	490 500 510 520	
5mdr1FUG.seq	GGCCTCCACCTTTATCACTTCTTTGTTATTGGTTTTGTA	489
3mdr1FUG.seq	CTTCGCCAACCTGGGCACCAGCATCATCATCAGCTTCGTG	504
		
	TATGGCTGGGAGCTGACCCTGGTCATCCTGGCCGTGGTGC	
	530 540 550 560	
5mdr1FUG.seq	CATGGATGGAAGCTCACCTGGTCATCCTGGCCATCAGCC	529
3mdr1FUG.seq	TACGGCTGGGAGCTGACCCTGCTCATCCTGGCCGTGGTGC	544
		
	CTGTGTTGGGTGTGTCCGGTGCCGTTTAGGTTAAGTTGCT	
	570 580 590 600	
5mdr1FUG.seq	CTGTGTTGGGTCTCTCAGCTGCCCTTACAGTAAGTTGCT	569
3mdr1FUG.seq	CCATCCTGGCTGTGGCCGGAGCCGCTGAGGTCAAGCTGCT	584
		
	GACAGGCTTCGCCGGTGAAGAGCAGACGGCGTTGGCGATG	
	610 620 630 640	
5mdr1FUG.seq	GACAAGCTTCACCAGTAAAGAGCAGACAGCGTACGCCAAA	609
3mdr1FUG.seq	GACAGGACACGCCGCGAAGACAAGAAGGAGCTGGAGATG	624
		
	GCTGGAGCTGTGGCCGCGGAGGTGCTCTCGTCTGTGAGGA	
	650 660 670 680	
5mdr1FUG.seq	GCTGGAGCTGTGGCAGCGGAGGTGCTATCCTCCATCAGGA	649
3mdr1FUG.seq	GCCGGAAAGATCGCCACAGAGGCCATCGAGAATGTGAGAA	664
		
	CTGTGTTTTCTTCAGTGGCGAACGGACATTTGTGGCTTT	
	690 700 710 720	
5mdr1FUG.seq	CTGTGTTTGCCTTCAGTGGCCAAAGAAAAGCCATCAAAG	689
3mdr1FUG.seq	CTGTGGTGTCCCTCACAGAGAACCGACATTTGTGGCTTT	704




 ATATGATGAGAATCTGGCTGTTGCGTGGGAGATGTGCCTG

 730 740 750 760

5mdr1FUG.seq ATATCATAAGAACCTGGAGGATGCGAGGGACATGGGAATA 729

 3mdr1FUG.seq ATACGAGGAAAATCTAACTGTTCCATACAAGAACTCCCAG 744

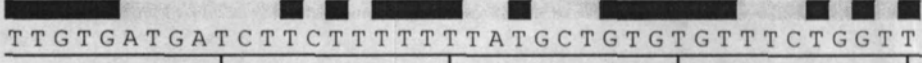


 AAGAAGGGCGTTGTTTCTGGCTTGGCCTCCTGCTTCTCCT

 770 780 790 800

5mdr1FUG.seq AAGAAGGGGAGTTGCTGCTAACACGGCCACAGGCTTCTCCT 769

 3mdr1FUG.seq AAAAAGGCCAAAATTTATGGCTTAACTACTCCTTCTCAC 784

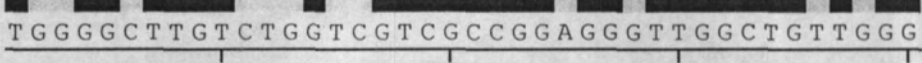


 TTGTGATGATCTTCTTTTTTTTATGCTGTGTGTTTCTGGTT

 810 820 830 840

5mdr1FUG.seq TTCTGATGATCTACCTGTCCTATGCTCTGGCCTTCTGGTA 809

 3mdr1FUG.seq AGGCCATGATCTTCTTTGTTTACGCTGCCTGTTTCCGCTT 824

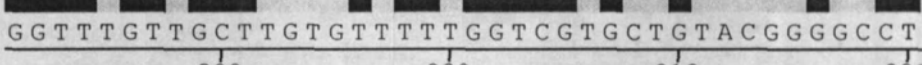


 TGGGGCTTGTCTGGTCGTCGCCGGAGGGTTGGCTGTTGGG

 850 860 870 880

5mdr1FUG.seq CGGGACTACTCTGGTCCTCAACAAAGAGTACACCATTGGA 849

 3mdr1FUG.seq TGGAGCCTGGCTGATCGAAGCAGGACGGATGGATGTGGAG 864

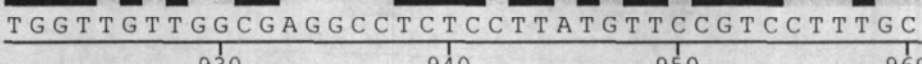


 GTTTTGTTGCTTGTGTTTTTGGTTCGTGCTGTACGGGGCCT

 890 900 910 920

5mdr1FUG.seq AATTTACTGACTGTGTTCTTCGTCGTCCTCTACGGGGCAT 889

 3mdr1FUG.seq GGAGTGTTCCCTTGTGGTTATGACAATGCTGTACGGCGCCA 904

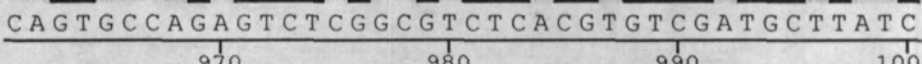


 TGGTTGTTGGCGAGGCCCTCTCCTTATGTTCCGTCCTTTGC

 930 940 950 960

5mdr1FUG.seq ACATTATTGGACAGGCCCTCTCCCAACGTCCAGTCCTTTGC 929

 3mdr1FUG.seq TGGCTGTCGGCGAGGCCAACACTTATGCTCCCAACTTCGC 944

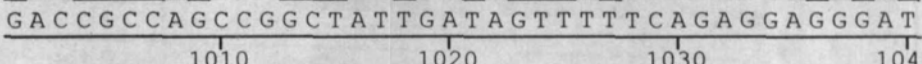


 CAGTGCCAGAGTCTCGGCGTCTCACGTGTCGATGCTTATC

 970 980 990 1000

5mdr1FUG.seq CAGTGCCAGAGGAGCGGCGTATAAAGTCTACAACATTATC 969

 3mdr1FUG.seq CAAAGCCAAAATCTCAGCCTCCCACTGACGATGCTAATA 984

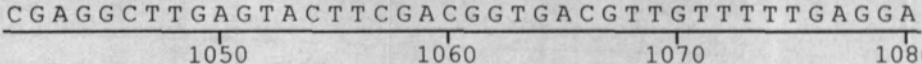


 GACCGCCAGCCGGCTATTGATAGTTTTTCAGAGGAGGGAT

 1010 1020 1030 1040

5mdr1FUG.seq GACCACAAACCGAATATTGACAGCTTTTCAGAGGACGGAT 1009

 3mdr1FUG.seq AACAGACAGCCGGCCATAGATAATCTGTCAGAGGAGGAAG 1024

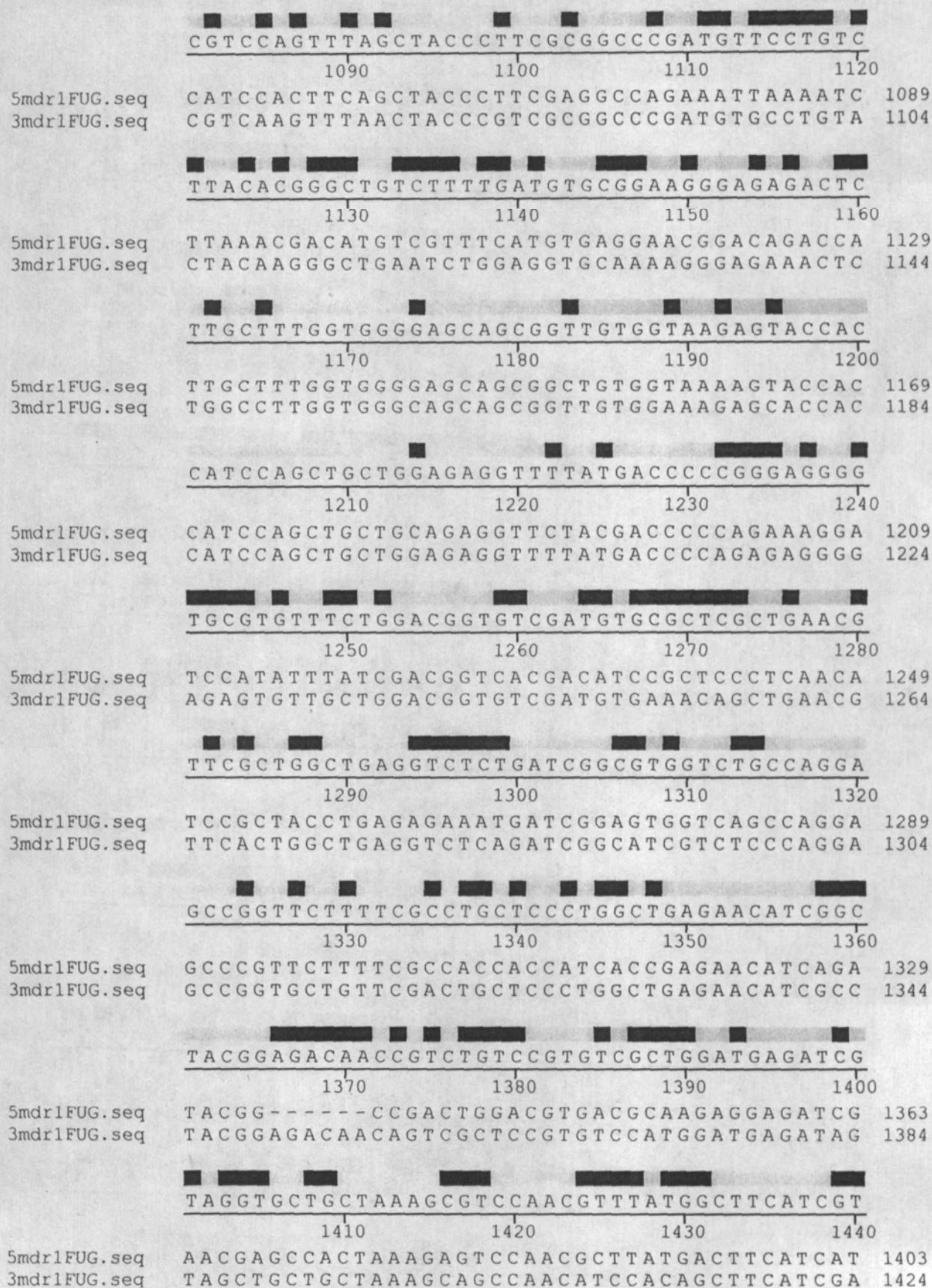


 CGAGGCTTGAAGTACTTCGACGGTGACGTTGTTTTTGAGGA

 1050 1060 1070 1080

5mdr1FUG.seq ACAAGCCTGAATACATCAAAGGTGACATTGTATTCCAGAA 1049

 3mdr1FUG.seq CGAGACTGGAGAAATACGACGGCAACGTTCTTTTTTGAGGA 1064



GGGGCTTCCTGAGAGGTTTGAGACTCTGGTTGGTGATCGG
1450 1460 1470 1480
5mdr1FUG.seq GAACCTTCCAGACAAGTTTGGAGACGCTGGTGGGAGATCGA 1443
3mdr1FUG.seq AGGGCTGCCTCAGAGATACGACACTCAGGCTGGTGATAAG 1464

GGGACTCAGCTGTGCGGGGGCCAGAAGCAGCGTGTGCGCA
1490 1500 1510 1520
5mdr1FUG.seq GGGACTCAGCTGAGCGGAGGACAGAAGCAGAGGATCGCCA 1483
3mdr1FUG.seq GGAACACAGCTGTGAGGGGGCCAGAAGCAGCGTGTGCGCA 1504

TCGCTCGAGCTCTGGTCCGCAACCCTAAACTGTTGCTGCT
1530 1540 1550 1560
5mdr1FUG.seq TCGCTCGAGCTCTGGTCCGCAACCCTAAAATCCTCCTGCT 1523
3mdr1FUG.seq TAGCCCGAGCCATCATCCGCAACCCCAAACCTGTTGCTCCT 1544

GGACGAGGCCACGTCTGCGCTCGATGCTGAGAGTGAGACG
1570 1580 1590 1600
5mdr1FUG.seq GGACGAGGCCACGTCTGCCCTCGATGCTGAGAGCGAGACC 1563
3mdr1FUG.seq GGACGAGGCCACGTCTGCGCTCGACACTGAGAGTGAGAAG 1584


GTGGTGCAGGCTGCTTTGGACCAGGTCCGGCTGGGTGCGGA
1610 1620 1630 1640
5mdr1FUG.seq ATCGTACAGGCTGCTCTGGACAAGGTCCGACTGGGTGCGCA 1603
3mdr1FUG.seq GTGGTGCAGGAGGCGTTGGACCAGGCCAGGAAGGGCAGGA 1624


CGTGCATCGTTCGCTCACCCTGCTGTCGACCATCCGGAA
1650 1660 1670 1680
5mdr1FUG.seq CCACCATCGTTCATCGCTCACCCTGCTGTCGACCATCAGAAA 1643
3mdr1FUG.seq CGTGCATCGTTCGTCGAGCCACCCTGCTGTCGACCATCCAGAA 1664

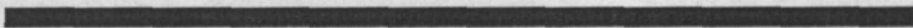
CGCCGACTTCATTGCTGTGTTCCGTTGGTGGTGTGGTGGTT
1690 1700 1710 1720
5mdr1FUG.seq CGCCGACATCATTGCTGGATTTCAGTAATGGTGAATCGTT 1683
3mdr1FUG.seq CGCCGACTGCATCGCTGTGTTCCAGGGAGGAGTGGTGGTG 1704


GAGCAGGGGACTCACCGGCAGCTGATGGCGATGAAGGGAG
1730 1740 1750 1760
5mdr1FUG.seq GAGCAGGGGACTCACAGCCAGCTGATGGAGATAAAGGGAG 1723
3mdr1FUG.seq GAAAAGGGGACGCACCAGCAGCTGATCGCCAAGAAGGGAG 1744


TGTATCATGTGCTGGTGACCATGCAGAGCTTTCAGAAGCT
1770 1780 1790 1800
5mdr1FUG.seq TCTATCATGGCCTGGTGACCATGCAGAGCTTTCAGAAGCT 1763
3mdr1FUG.seq TGTACCACATGCTGGTGACCAAACAGA----- 1772


 GGAGGATCTGGAAGACTCAGACTACGAGCCCTGGGTTCGCT
 1810 1820 1830 1840
 5mdr1FUG.seq GGAGGATCTGGAAGACTCAGACTACGAGCCCTGGGTTCGCT 1803
 3mdr1FUG.seq ----- 1772




 GAGAAGAGCCAGCTGATCGAATCCTTCTCCCAGTCCTCCC
 1850 1860 1870 1880
 5mdr1FUG.seq GAGAAGAGCCAGCTGATCGAATCCTTCTCCCAGTCCTCCC 1843
 3mdr1FUG.seq ----- 1772


 TGCAGAGGAGGAGGTCCACTAGAGGCTCCTTGCTTGCTGT
 1890 1900 1910 1920
 5mdr1FUG.seq TGCAGAGGAGGAGGTCCACTAGAGGCTCCTTGCTTGCTGT 1883
 3mdr1FUG.seq ----- 1772


 CTCAGAAGGAACAAAAGAGGAGAAAGAAAAATTTGAGTGC
 1930 1940 1950 1960
 5mdr1FUG.seq CTCAGAAGGAACAAAAGAGGAGAAAGAAAAATTTGAGTGC 1923
 3mdr1FUG.seq ----- 1772


 GATTGGGCTATCATAGTGGAX
 1970 1980
 5mdr1FUG.seq GATCAGGACAACATAGAGGAG 1944
 3mdr1FUG.seq ---TGGGCTATCACAGTGA 1788

Percent Similarity

	1	2		
1		52.6	1	5mdr1FUG.seq
2	75.7		2	3mdr1FUG.seq
	1	2		

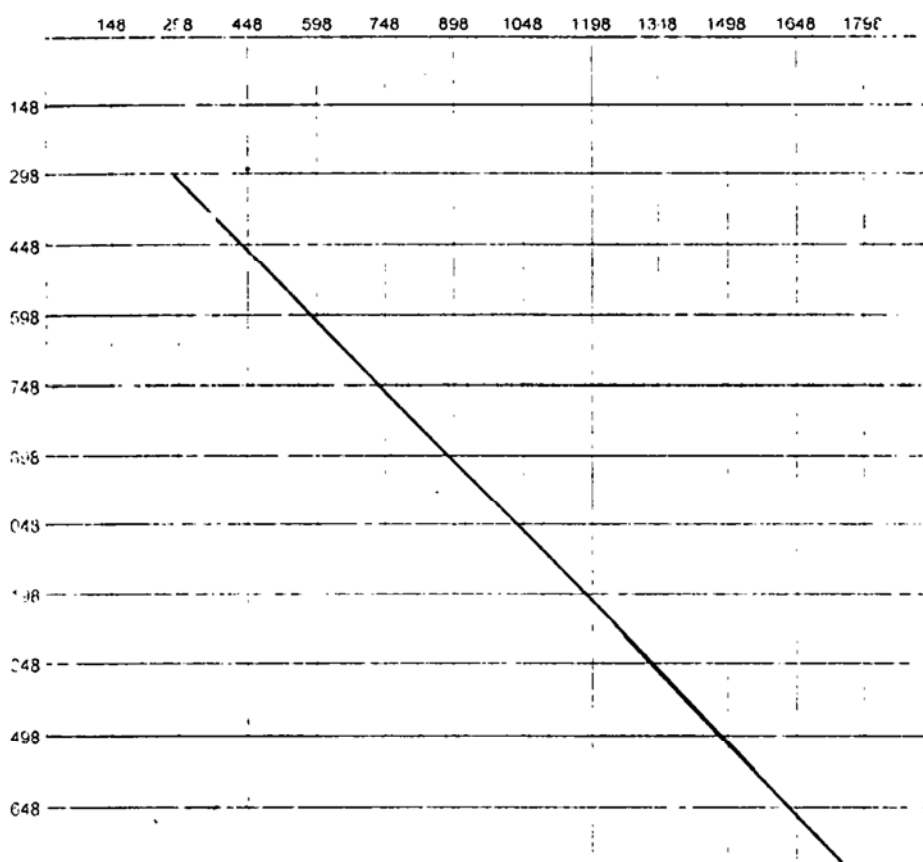
fuguMDR1 cDNA 的 N 端半侧和 C 端半侧的同源性比较结果:

	DENVPPVSFFKVF	FAVSGWDILLVTIGVLC	AVNGAVN	
	10	20	30	40
5mdr1FUG.seq	-----FRFA-	DGWDILMVTIGVLM	AI V NGLVNP	79
3mdr1FUG.seq	DENVPPVSFFKVM	RYNVSEWPYILV--	GTICAMINGAMQP	112
	VFSIVFGEIIDVF	IEEAKLIINENS	SNPTSTMSNVL	RVCC
	50	60	70	80
5mdr1FUG.seq	LMCIVFGEMTDS	FIQEAKLSQNHNT	SNPTSTMSNVL	RVCC 199
3mdr1FUG.seq	VFSIIFTEIIMV	FREKDKEIIREKSS	-----	193
	IIRHLFSILFAVL	GVAVLVVAF	LQMSLWGLQG	FCSKSGA
	90	100	110	120
5mdr1FUG.seq	IIRHLFSIYYSI	LGFVAVLVVAYL	QMSLWTL-----	TAA 298
3mdr1FUG.seq	----FFCILFAV	MGVVTFLTMLQ--	FWGFQGFCSKSGE	292
	ILALNLRLLAFI	GIMQQDLSWYD	VTETT	VGALTTRLAADA
	130	140	150	160
5mdr1FUG.seq	RQAKRIRELFFH	GIMQQDISWYD	TET--GELNTRLTE--	409
3mdr1FUG.seq	ILTLNLRRLKAF	ISMQRDLSWYD	NPKN TVGALTTRLAADA	412
	AHVQGGAGVKAG	VLIQAAATL	LGTSIVIGFV	HGWELTLVIL
	170	180	190	200
5mdr1FUG.seq	---WEGIGDKAG	LLIQAASTFIT	SFVIGFVHGWK	LTLVIL 517
3mdr1FUG.seq	AHVQGAAGVRLA	VMTQNFANL	GTSIIISFVY	GWELTLIL 532
	AVVPVLGVAGAA	EVKLLTGHA	AEDQTALAKA	GAVAAEALS
	210	220	230	240
5mdr1FUG.seq	AISPVLGLSAA	LYSKLLTSF	TKEQTAYAKA	GAVAAEVLS 637
3mdr1FUG.seq	AVVPILAVAGAA	EVKLLTGHA	AEDKKELEMAG	KIATEAIE 652
	SVRTVVALSGER	TAVALYEENL	TVARDNGIKK	GVAAGLAT
	250	260	270	280
5mdr1FUG.seq	SIRTVFAFSGQ	RKAIKRYHKN	LEDARDMGIKK	GVAANTAT 757
3mdr1FUG.seq	NVRTVVSALTRE	PTFVALYEENL	TPYKNSQKKAK	IYGLTY 772
	GFSQAMIFLVYAA	AFRFGATLVL	AGEYTVGGV	L LVFVVL
	290	300	310	320
5mdr1FUG.seq	GFSFLMIYLSY	ALAFWYGTT	LVLNKEYTIGN	LLTVFFVVL 877
3mdr1FUG.seq	SFSQAMIFFVYA	ACFRFGAWL	IEAGRMDVEG	VFLVMTML 892
	YGAYAVGEASTN	AQSFA	SAKGAASKV	TNLI DHQPAIDSL
	330	340	350	360
5mdr1FUG.seq	YGAYIIGQAS	PNVQSFASARG	AAYKVYNIID	HKPNIDSFS 997
3mdr1FUG.seq	YGAMAVGEANTY	APNFAKAKIS	ASHLTMLINR	QPAIDNLS 1012

	EDGAKLEKIDGDVVFEDVKFSYPSRPDVKVLNGLSLEVQN	
	370 380 390 400	
5mdr1FUG.seq	EDGYKPEYIKGDIVFQNIHFSYPSRPEIKILNDMSFHVRN	1117
3mdr1FUG.seq	EEEARLEKYDGNVLFEDVKFNYPSPRPDVPVLQGLNLEVQK	1132
	GETLALVGSSGCGKSTTIQLLERFYDPQEGSVLLDGVDVK	
	410 420 430 440	
5mdr1FUG.seq	GQTIALVGSSGCGKSTTIQLLQRFYDPQKGSIFIDGHDIR	1237
3mdr1FUG.seq	GETLALVGSSGCGKSTTIQLLERFYDPREGRVLLDGVDVK	1252
	SLNVHYLRSQIGVVSQEPVLFATSLAENIAYGDNSLSVSQ	
	450 460 470 480	
5mdr1FUG.seq	SLNIRYLREMIGVVSQEPVLFATTITENIRYG--RLDVTQ	1351
3mdr1FUG.seq	QLNVHWLRSQIGIVVSQEPVLFDCSLAENIAYGDNSRSVSM	1372
	DEIVAAAKAANAHSFIEGLPDKFDTLAGDKGTQLSGGQKQ	
	490 500 510 520	
5mdr1FUG.seq	EEIERATKESNAYDFIMNLPDKFETLVGDRGTQLSGGQKQ	1471
3mdr1FUG.seq	DEIVAAAKAANIHSFIEGLPQRYDTQAGDKGTQLSGGQKQ	1492
	RVAIARALVRNPKLLLLDEATSALDAESETVVAALDQAR	
	530 540 550 560	
5mdr1FUG.seq	RIAIARALVRNPKILLLLDEATSALDAESETIVQAALDKVR	1591
3mdr1FUG.seq	RVAIARAIRNPKLLLLDEATSALDTESEKVVQEALDQAR	1612
	LGRTTIVVAHRLSTIQNADI IAGFSGGVVVEQGTHSQLIA	
	570 580 590 600	
5mdr1FUG.seq	LGRTTIVIAHRLSTIRNADI IAGFSNGEIVEQGTHSQLME	1711
3mdr1FUG.seq	KGRTCIVVAHRLSTIQNADCI AVFQGGVVVEKQTHQQLIA	1732
	IKGVYHGLVTKQSFQKLEDLEDSEWVAEKSQ LIESFS	
	610 620 630 640	
5mdr1FUG.seq	IKGVYHGLVTKQSFQKLEDLEDSDYEPWVAEKSQ LIESFS	1831
3mdr1FUG.seq	KKGVYHMLVTKQM-----GYHS-----	1786
	QSSLQRRRSTRGSL LAVSEGTKEEKEKFEC DQDNIEG	
	650 660 670	
5mdr1FUG.seq	QSSLQRRRSTRGSL LAVSEGTKEEKEKFEC DQDNIEE	1942
3mdr1FUG.seq	-----G	1786

Percent Similarity				
	1	2		
1	■	37.8	1	5mdr1FUG.seq
2	100.0	■	2	3mdr1FUG.seq
	1	2		

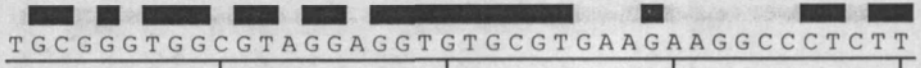
Dotplot 分析结果: fuguMDR1 N 端为横坐标; C 端为纵坐标



FuguMDR2 cDNA 的 5'端半侧和 3'端半侧的同源性比较结果:

	TTCCGATTTGGAGGTTGCTGGGGTGTTCGTTCTTTTCATTG	
	10 20 30 40	
5mdr2FUG.seq	TTCCGATTTGCAGACCGCTGGGACGTCGTGCTGCTCATCA	40
3mdr2FUG.seq	G-----ATGAAGACGTTCCCGACGTGTTCGTTCTTTAAAGTG	36
	GTGGGTCTGTGCATGGCCGTGTTGCCCTGCATCGTTGTGG	
	50 60 70 80	
5mdr2FUG.seq	GCGGGACAGTG-ATGGCCATGGTCAACGGCACCGTGATGC	79
3mdr2FUG.seq	CTGCATCTGAACATCCCCGAGTGGCCCTACATCCTTGTGG	76
	GGCTCATGTGCGTTGTGTTTGGTGGAGTGATGGAGCGTTT	
	90 100 110 120	
5mdr2FUG.seq	CCCTCATGTGCATTGTCTTTGGAGAAATGACGGACAGTTT	119
3mdr2FUG.seq	GGCTCATCTGCGCTACGATCAATGGAGCCATGCAGCCGGT	116
	TTTCTCCGTTGTCTTGTCCCAGCTCATCGCTAGTGTTTGG	
	130 140 150 160	
5mdr2FUG.seq	TATATACGCTGACATGGCCCAACACAACGCAAGTGGCTGG	159
3mdr2FUG.seq	CTTCGCCATCCTCTTCTCCAAGATCATCACT-GTGTTTGC	155
	GGTTCTAGTACTGTTATTCTGTACAGCACGTTACAGGAGG	
	170 180 190 200	
5mdr2FUG.seq	AATTCTACTACTACTATTCTGAACAGCACGTTACAGGAGG	199
3mdr2FUG.seq	GGATCCAG-ACCGTGATTCTGT-----CAGGAGG	183
	ACGTGTGAATTTTTTTCTTTTATTTTTGTCGTTTTTGGCT	
	210 220 230 240	
5mdr2FUG.seq	ACATGCAAAGATTCGCCATTTATTACT-CCGTCTTGGGAT	238
3mdr2FUG.seq	AAGAGTGAATTCATTTCTCTGATGTTTGTTCGTTATTGGCT	223
	TTGTTTTGTTTGTGGCCGTGTTTTTGCAGGTTTCCTTTTT	
	250 260 270 280	
5mdr2FUG.seq	TTGTTGTGCTGCTGGCCGCCTACATGCAGGTGTCCTTCTG	278
3mdr2FUG.seq	GTGTGTCATTTGTCAACCATGTTTTTACAGGGTTACTGTTT	263
	GGGTATATCCGGCGGGCTTCTGGTGCTGCGGCTGCGCCTC	
	290 300 310 320	
5mdr2FUG.seq	GACCATAACAGCCGGGCGCCAGGTGAAACGCATCCGCAGC	318
3mdr2FUG.seq	CGGTAAATCCGGAGAGATTCTGACGCTGAAGCTGAGACTC	303
	TTGTTTTTCCCGTGGATGATGCGGCAGGAGCTCAGCTGGT	
	330 340 350 360	
5mdr2FUG.seq	TTGTTTTTCCACTGCATCATGCAGCAGGAGATCAGCTGGT	358
3mdr2FUG.seq	CGGGCGTTTACGGCGATGATGAGACAGGACCTCAGCTGGT	343

	TTGACGTTCCCCAGAACGCCGTTGGGGCTCTCACCCTCG	
	370 380 390 400	
5mdr2FUG.seq	TTGACGT-----GAACGACACAGGGGAGCTCAACACTCG	392
3mdr2FUG.seq	ACGACAATCCCCAAAACACCGTTGGCGCTCTCACCCTAG	383
	GCTGGCTGCCGACGTCTCCCAGGTCAGGGGGGTGTCCGGT	
	410 420 430 440	
5mdr2FUG.seq	ACTGACGGACGACGTCTACAAGATCCAGGAGGGCATCGGT	432
3mdr2FUG.seq	GCTGGCTGCCGACGCCGCCAAGTACAAGGAGCTGCAGGG	423
	GTGCGGGTGGGGCTGCTGATGCAGGCGTTTCGCCACCTTGG	
	450 460 470 480	
5mdr2FUG.seq	GACAAGGTGGGGCTGCTGATCCAGGCGTACACCACCTTCA	472
3mdr2FUG.seq	GTGCGCCTGGCGACAATAATGCAGAACTTCGCCAACCTGG	463
	TCACGGGCTTCATCATCGGCTTTGTTTCGGGCTGGGAGCT	
	490 500 510 520	
5mdr2FUG.seq	TCACGGCCTTCATCATCGGCTTCACCACGGGCTGGAACT	512
3mdr2FUG.seq	GCACCAGCATCATCATCGCCTTTGTTTACGGCTGGGAGCT	503
	GACGCTGGTTCATCCTGGCCGTGGTGCCCGTGCTGGCGGTT	
	530 540 550 560	
5mdr2FUG.seq	GACGCTGGTTCATCCTGGCCGTGAGCCCCGCGCTGGCCATT	552
3mdr2FUG.seq	GACCCTGCTCATCCTGGCCGTGGTGCCCTCATCGCGGCC	543
	TCGGGCGCCTTTTTGATTAAGGTGCTTGCGTGTTCGCCG	
	570 580 590 600	
5mdr2FUG.seq	TCGGCCGCCTTCTTCAGTAAAGTGCTTGCGTCTTCACCA	592
3mdr2FUG.seq	GCCGGAGCCGCTGAGATCAAGCTGCTCGCGGGTCACGCCG	583
	GTAAGGAGCAGACGGCGTTGGCGAAGGCCGGAGCGGTGGC	
	610 620 630 640	
5mdr2FUG.seq	GTAAGGAGCAGACGGCGTACGCCAAAGCCGGAGCCGTGGC	632
3mdr2FUG.seq	CAAAGACAAGAAGGAGCTGGAGAAGGCCGGAAAGATCGC	623
	GGCGGAGGTGCTGTTCGGCCGTACAGGACCGTGTTCCTTC	
	650 660 670 680	
5mdr2FUG.seq	GGAGGAAGTGCTGTCCGCCATCAGGACCGTGTTCGCCTTC	672
3mdr2FUG.seq	CACAGAGGCCATCGAGAACGTCAGAACCCTCGTGTCCCTC	663
	AGTGGTGAGCCCAGATTTGTTTGTATATGAGGAGAATC	
	690 700 710 720	
5mdr2FUG.seq	AGTGGTCAGACCAGAGAGATTGAGAGATACCACAAGAACC	712
3mdr2FUG.seq	AGCAGAGAACCAAAATTTGAGTGTATATGAGGAGAATC	703

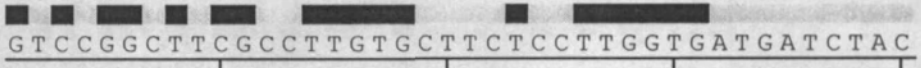


 TGCGGGTGGCGTAGGAGGTGTGCGTGAAGAAGGCCCTCTT

 730 740 750 760

5mdr2FUG.seq TGCGGGACGCAAAGGACGTGGGAGTGAAGAAGGCCATCTC 752

 3mdr2FUG.seq TCAGAGTGCCGTACAAGAACTCCAGAAAAAGGCCACGT 743

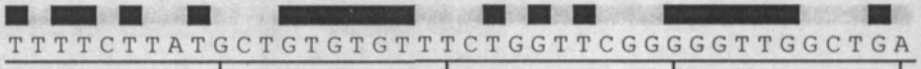


 GTCCGGCTTCGCCTTGTGCTTCTCCTTGGTGATGATCTAC

 770 780 790 800

5mdr2FUG.seq CTCCAACATCGCCATGGGCTTTCACCTTCTGATGATCTAC 792

 3mdr2FUG.seq GTACGGCTTAACCTACTCCTTCTCCAGGCCATGATCTAC 783

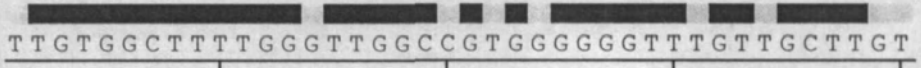


 TTTTCTTATGCTGTGTGTTTCTGGTTCGGGGGTTGGCTGA

 810 820 830 840

5mdr2FUG.seq CTGTCCTATGCTCTGGCCTTCTGGTACGGGAGTACGCTCA 832

 3mdr2FUG.seq TTTGCTTACGCTGCCTGTTTCCGCTTCGGAGCCTGGCTGA 823

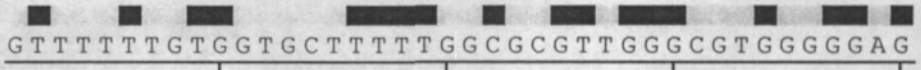


 TTGTGGCTTTTGGGTTGGCCGTGGGGGTTTGTGTTGCTTGT

 850 860 870 880

5mdr2FUG.seq TCCTGAATTTTGGAGTACACCATCGGCAATTTACTGACTGT 872

 3mdr2FUG.seq TTGAAGCAGGACGGATGGACGTGGAGGGAGTGTTCTGTT 863

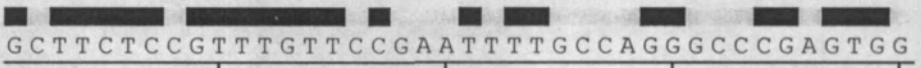


 GTTTTTTGTGGTGCTTTTTTGGCGCGTTGGGCGTGGGGGAG

 890 900 910 920

5mdr2FUG.seq GTTTTTTGTGCGTGCTTATTGGAGCGTTCAGCGTCGGACAG 912

 3mdr2FUG.seq GGTTTTCTGCGGTGCTGTACGGCGCCATGGCCGTGGGGGAA 903

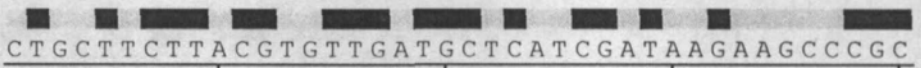


 GCTTCTCCGTTTGTTCGGAATTTTGCCAGGGCCCGAGTGG

 930 940 950 960

5mdr2FUG.seq ACCTCTCCGAACATCCAGAATTTTGCCAGCGCCCGAGGAG 952

 3mdr2FUG.seq GCTAACACCTTTGCTCCGAACTACGCCAAGGCCAAAATGG 943




 CTGCTTCTTACGTGTTGATGCTCATCGATAAGAAGCCCGC

 970 980 990 1000

5mdr2FUG.seq CCGCCTATAAAGTGTACAGCATCATCGATAACAAGCCAAA 992

 3mdr2FUG.seq CTGCTTCTTACCTGATGATGCTAATAACAAGAAGCCCGC 983

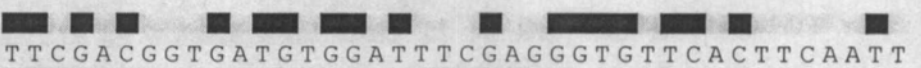


 CATTGATAGCTTTTTCAGAGGAGGGTTTGTCTCCGGACTTC

 1010 1020 1030 1040

5mdr2FUG.seq CATTGACAGCTTTTTCAGAGGACGGTTTCAAGCCGGACTTC 1032

 3mdr2FUG.seq CATTGATAACCTCTCAGAGGAGGGGACGTCTCCGGAAAA 1023

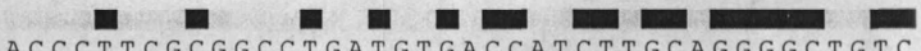
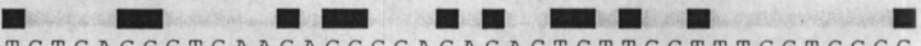
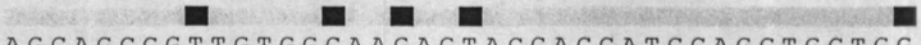
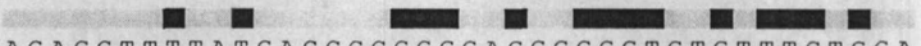
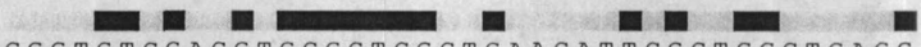
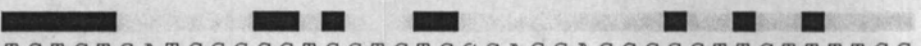
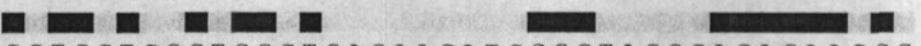
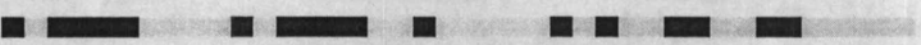
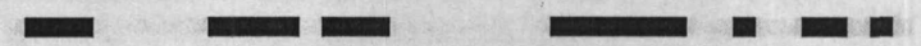


 TTCGACGGTGATGTGGATTTTCGAGGGTGTTCACTTCAATT


 1050 1060 1070 1080


5mdr2FUG.seq ATCAAAGGTGACATCGAGTTCAAGAACATCCACTTCAATT 1072


 3mdr2FUG.seq TACGACGGTAATGTGCATTTTCGAGGGTGTAAATTCAACT 1063


		
	ACCCTTCGCGGCCTGATGTGACCATCTTGCAGGGGCTGTC	
	1090 1100 1110 1120	
5mdr2FUG.seq	ACCCTTCGAGGCCTGAAGTCAAAATCTTGAACAACATGTC	1112
3mdr2FUG.seq	ACCCGTCGCGGCCCGATGTGACCATACTCCAGGGGCTGAA	1103
		
	TCTGAGGGTGAAGAGGGGAGAGACTCTTGCTTTGGTGGGG	
	1130 1140 1150 1160	
5mdr2FUG.seq	TCTGAGCGTGAAGAGCGGACAGACCATTGCTTTGGTGGGG	1152
3mdr2FUG.seq	CCTGAAGGTGAAAAGGGAGAACTCTGGCCTTGGTGGGC	1143
		
	AGCAGCGGTTGTGGCAAGAGTACCACCATCCAGCTGCTGG	
	1170 1180 1190 1200	
5mdr2FUG.seq	AGCAGCGGCTGTGGCAAAGTACCACCATCCAGCTGCTGC	1192
3mdr2FUG.seq	AGCAGCGGTTGTGGAAAGAGCACCACCATCCAGCTGCTGG	1183
		
	AGAGGTTTTATGACCCCGGGGAGGGGGTGTGTTTTCTGGA	
	1210 1220 1230 1240	
5mdr2FUG.seq	AGAGGTTCTACGACCCCGAGGAAGGAGCTGTATTTATCGA	1232
3mdr2FUG.seq	AGAGGTTTTATGACCCCGAGAGAGGGGAGAGTGTCACTGGA	1223
		
	CGGTGTCGACGTGCGCTCGCTGAACATTCGCTGGCTGAGG	
	1250 1260 1270 1280	
5mdr2FUG.seq	CGGTCACGACATCCGCTCCCTCAACATCCGCTACCTGAGA	1272
3mdr2FUG.seq	CGGTGTC AACGTGAAACAGCTGAACATTC ACTGGCTGAGG	1263
		
	TCTCTGATCGGCGTGGTCTGCCAGGAGCCGGTTCTTTTCG	
	1290 1300 1310 1320	
5mdr2FUG.seq	GAGATGATCGGAGTGGTCCAGCCAGGAGCCCGTTCTTTTCG	1312
3mdr2FUG.seq	TCTCAGATCGGCATCGTCTCCAGGAGCCGGTGCTGTTCG	1303
		
	CCTGCTCCCTGGCTGAGAACATCGGCTACGGAGACAACCG	
	1330 1340 1350 1360	
5mdr2FUG.seq	CCACCACCATCACCGAGAACATCAGATACGG-----CCG	1346
3mdr2FUG.seq	ACTGCTCCCTGGCTGAGAACATCGCCTACGGAGACAACAG	1343
		
	TCTGTCCGTGTCGCTGGATGAGATCGTACTTGCTGCTAAA	
	1370 1380 1390 1400	
5mdr2FUG.seq	ACTGGACGTGACGCAAGAGGAGATCGAACGAGCCACTAAA	1386
3mdr2FUG.seq	TCGCTCCGTGTCATGGATGAGATAGTACTTGCTGCTAAA	1383
		
	GCGTCCAACGTTTTATGGCTTCATCGTGGGGCTTCCTGAGA	
	1410 1420 1430 1440	
5mdr2FUG.seq	GAGTCCAACGCTTATGACTTCATCATGAACCTTCCAGACA	1426
3mdr2FUG.seq	GCAGCCAACATCCACAGCTTCATCGAAGGGCTGCCTCAGA	1423

	GGTTTGAGACTCTGGTTGGTGATCGGGGGACTCAGCTGTG	
	1450 1460 1470 1480	
5mdr2FUG.seq	AGTTTGAGACGCTGGTGGGAGATCGAGGGACTCAGCTGAG	1466
3mdr2FUG.seq	GATACGACACTCAGGCTGGTGATAAGGGAACACAGCTGTC	1463
	CGGGGGCCAGAAGCAGCGTGTGCGCCATCGCTCGAGCTCTG	
	1490 1500 1510 1520	
5mdr2FUG.seq	CGGAGGACAGAAGCAGAGGATCGCCATCGCTCGAGCTCTG	1506
3mdr2FUG.seq	AGGGGGCCAGAAGCAGCGTGTGCGCCATAGCCCGAGCCATC	1503
	GTCCGCAACCCTAAACTGTTGCTGCTGGACGAGGCCACGT	
	1530 1540 1550 1560	
5mdr2FUG.seq	GTCCGCAACCCTAAAATCCTCCTGCTGGACGAGGCCACGT	1546
3mdr2FUG.seq	ATCCGCAACCCCAAACCTGTTGCTCCTGGACGAGGCCACGT	1543
	CTGCGCTCGATGCTGAGAGTGAGACGGTGGTGCAGGCTGC	
	1570 1580 1590 1600	
5mdr2FUG.seq	CTGCCCTCGATGCTGAGAGCGAGACCATCGTACAGGCTGC	1586
3mdr2FUG.seq	CTGCGCTCGACACTGAGAGTGAGAAGGTGGTGCAGGAGGC	1583
	TTTGGACCAGGTCCGGCTGGGTCCGACGTGCATCGTGGTC	
	1610 1620 1630 1640	
5mdr2FUG.seq	TCTGGACAAGGTCCGACTGGGTCCGACCATCGTGGTGC	1626
3mdr2FUG.seq	GTTGGACCAGGCCAGGAAGGGCAGGACGTGCATCGTGGTA	1623
	GCTCACCGTCTGTGCGACCATCCGGAACGCCGACTTCATTG	
	1650 1660 1670 1680	
5mdr2FUG.seq	GCTCACCGACTCTCGACCATCAGAAACGCCGACATCATTG	1666
3mdr2FUG.seq	GCCACCGTCTGTCCACCATCCAGAACGCCGACTGCATCG	1663
	CTGTGTTCCGTGGTGGCGTGGTGGTGGAGCAGGGGACTCA	
	1690 1700 1710 1720	
5mdr2FUG.seq	CTGGATTTCAGTAATGGCAAATCGTGGAGCAGGGGACTCA	1706
3mdr2FUG.seq	CTGTGTTCCAGGGAGGAGTGGTGGTGGAAAAGGGGACGCA	1703
	CCGGCAGCTGATGGCGATGAAGGGAGTGTATCATGTGCTG	
	1730 1740 1750 1760	
5mdr2FUG.seq	CAGCCAGCTGATGGAGATAAAGGGAGTGTATCATGGCCTG	1746
3mdr2FUG.seq	CCAGCAGCTGATCGCCAAGAAGGGAGTGTACCACATGCTG	1743
	GTGACCATGCAGATGTTCTATCATGTGGAGGAGGAAAATA	
	1770 1780 1790 1800	
5mdr2FUG.seq	GTGACCATGCAGACGTTCCACAATGTGGAGGAGGAAAATA	1786
3mdr2FUG.seq	GTCACCAAACAGATGGGCTATCACA-----	1769




CCGCCATGTCGGAGTTATCTGCTGGGGAGAAGAGCCCTGT
 1810 1820 1830 1840
 5mdr2FUG.seq CCGCCATGTCGGAGTTATCTGCTGGGGAGAAGAGCCCTGT 1826
 3mdr2FUG.seq ----- 1769


GGAAAAGACCGTCTCCAGTCGTCCATCATCAGGAGGAAG
 1850 1860 1870 1880
 5mdr2FUG.seq GGAAAAGACCGTCTCCAGTCGTCCATCATCAGGAGGAAG 1866
 3mdr2FUG.seq ----- 1769


TCCACCAGAGGGTCCTCGTTTGCCGCGTCAGAAGGAACCA
 1890 1900 1910 1920
 5mdr2FUG.seq TCCACCAGAGGGTCCTCGTTTGCCGCGTCAGAAGGAACCA 1906
 3mdr2FUG.seq ----- 1769

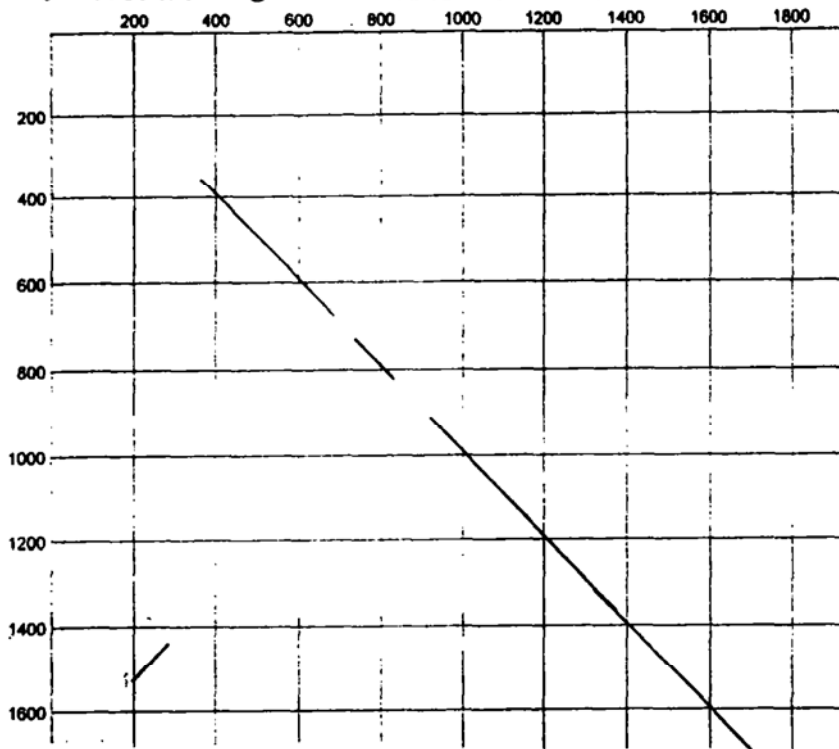

AAGAGGAAAAGACAGACGAG
 1930 1940
 5mdr2FUG.seq AAGAGGAAAAGACAGAAGAG 1926
 3mdr2FUG.seq -----ACGAC 1773

Percent Similarity

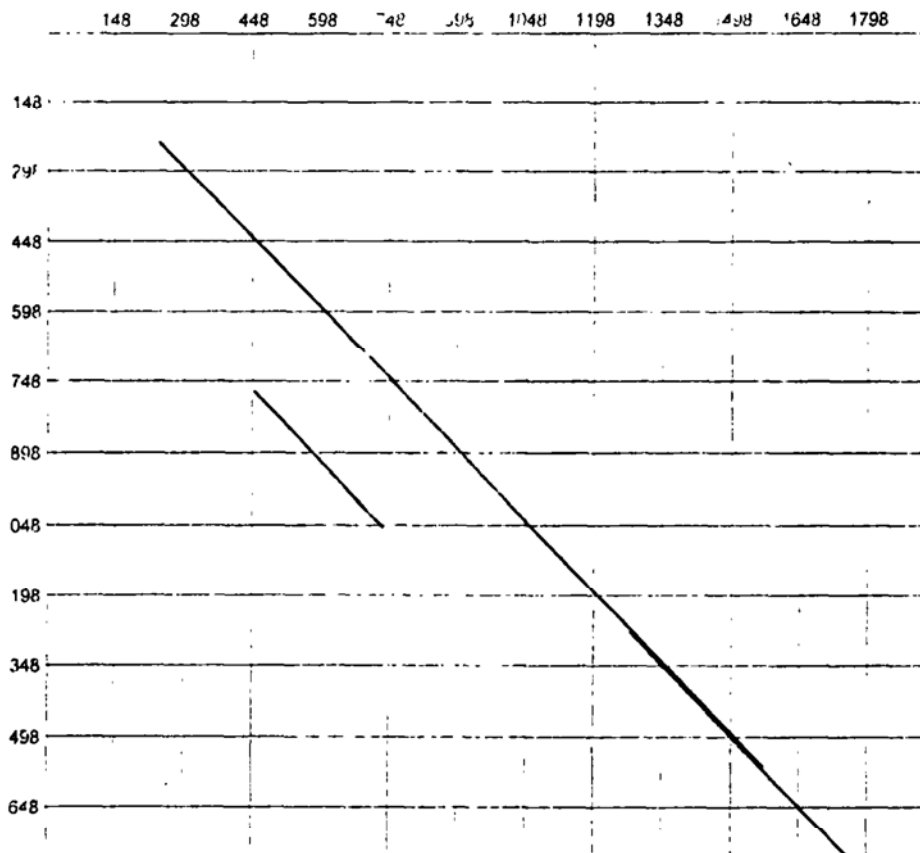
	1	2	
1		54.3	1
2	70.7		2
	1	2	

5mdr2FUG.seq
3mdr2FUG.seq

Dotplot 分析结果: fuguMDR2 5'端为横坐标, 3'端为纵坐标



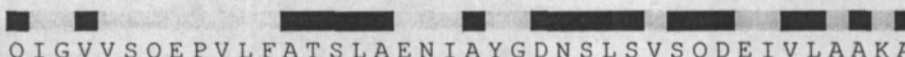
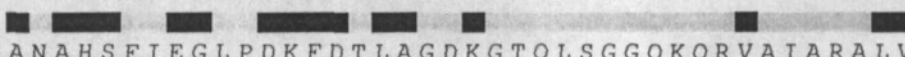
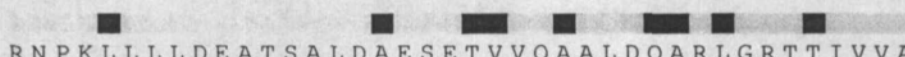
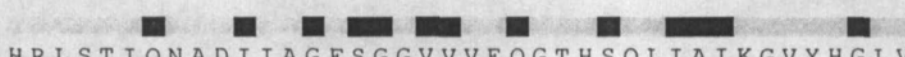
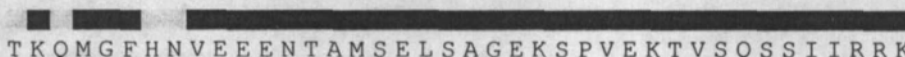
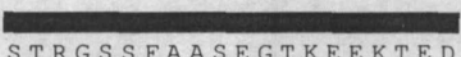


Dotplot 分析结果: fuguMDR2 N 端为横坐标, C 端为纵坐标



FuguMDR2 cDNA 的 N 端半侧和 C 端半侧的同源性比较结果:

	DEDVPDVSFFKVDLNIPEWVVLVISGLVCATVNGAVQP	
	10 20 30 40	
5MDR2FUG.SEQ	-----FRFAD-----RWDVLLISGTVMMAMVNGTVMP	79
3mdr2FUG.seq	DEDVPDVSFFKVLHLNIPEWPYILV--GLICATINGAMQP	112
	VFAIVFGEIITVFADADMAQHNASGWNSTTTILNSSVQED	
	50 60 70 80	
5MDR2FUG.SEQ	LMCIVFGEMTDSFIYADMAQHNASGWNSTTTILNSTLQED	199
3mdr2FUG.seq	VFAILFSKIITVFADPD-----RDSVRRK	184
	SERIALYFVVLGFVVLVA AFLQGSFFGISGGILVLKRLRL	
	90 100 110 120	
5MDR2FUG.SEQ	MQRFAIYYSVLGFVLLAAYMQVSFWTITAGRQVKRIRSL	319
3mdr2FUG.seq	SEFISLMFVVGCVS FVTMFLQGYCFGKSGEILTLKRLRL	304
	AFTAIMQQDLSWFDVPQNTVGALTTTLAADAQVQGGAGV	
	130 140 150 160	
5MDR2FUG.SEQ	FFHCIMQQEISWFDV--NDTGELNTRLTDDVYKIQEGIGD	433
3mdr2FUG.seq	AFTAMMRQDLSWYDNPQNTVGALTTTLAADAQVQGAAGV	424
	KVGLLIQAFATLGTAIIGFVTGWELTLVILAVVPALAAA	
	170 180 190 200	
5MDR2FUG.SEQ	KVGLLIQAYTTTITAFIIGFTTGWKLTLVILAVSPALAIS	553
3mdr2FUG.seq	RLATIMQNANLGTSSIIIAFVYGWELTLVILAVVPLIAAA	544
	GAAEIKVLAGHAAKDQTALAKAGAVATEALSVRTVVALS	
	210 220 230 240	
5MDR2FUG.SEQ	AFFFSKVLASFTSKEQTAYAKAGAVAEVLSAIRTVFAFS	673
3mdr2FUG.seq	GAAEIKLLAGHAAKDKKELEKAGKIATEAIENVRTVVSLS	664
	GETKEIELYEENLRVAKDVGKKAIVSGLAYGFSQAMIYL	
	250 260 270 280	
5MDR2FUG.SEQ	GQTREIERYHKNLRDAKDVGKKAISSNIAMGFTFLMIYL	793
3mdr2FUG.seq	REPKFECLYEENLRVPYKNSQKKAHVYGLTYSFSQAMIYF	784
	AYAAAFRFGATLILAGEYTVGGVLLVVS AVLIGAFVAVGEA	
	290 300 310 320	
5MDR2FUG.SEQ	SYALAFWYGSTLILNFEYTIGNLLTVFFVVLIGAFSVGQT	913
3mdr2FUG.seq	AYAACFRFGAWLIEAGRMDVEGVFLVVS AVLYGAMAVGEA	904
	STNAQNFASAKGAASKVYSLIDNKPAIDSLSEDGTS PDKI	
	330 340 350 360	
5MDR2FUG.SEQ	SPNIQNFA SARGAAYKVYSIIDNKPNIDSFSEDGFKPDFI	1033
3mdr2FUG.seq	NTFAPNYAKAKMAASYLMMLINKKPAIDNLSEEGTSPEKY	1024

		
	DGDVEFEGVKFNYPSPDVTILNGLSLSVKSGETLALVGS	
	370 380 390 400	
5MDR2FUG.SEQ	KGDIEFKNIHFNYPSPRPEVKILNNMSLSVKSGQTIALVGS	1153
3mdr2FUG.seq	DGNVHFEGVKFNYPSPDVTILQGLNLKVKKGETLALVGS	1144
		
	SGCGKSTTIQLLERFYDPEEGAVSLDGVVDVKSLSNIHYLRS	
	410 420 430 440	
5MDR2FUG.SEQ	SGCGKSTTIQLLQRFYDPEEGAVFIDGHDIRSLNIRYLRE	1273
3mdr2FUG.seq	SGCGKSTTIQLLERFYDPREGRVSLDGVNVKQLNIHWLRS	1264
		
	QIGVVSQEPVLFATSLAENIAYGDNSLSVSQDEIVLAAKA	
	450 460 470 480	
5MDR2FUG.SEQ	MIGVVSQEPVLFATTITENIRYG--RLDVTQEEIERATKE	1387
3mdr2FUG.seq	QIGIVSQEPVLFDCSLAENIAYGDNSRSVSMDEIVLAAKA	1384
		
	ANAHSFIEGLPDKFDTLAGDKGTQLSGGQKQRVAIARALV	
	490 500 510 520	
5MDR2FUG.SEQ	SNAYDFIMNLPDKFETLVGDRGTQLSGGQKQRIAIARALV	1507
3mdr2FUG.seq	ANIHSFIEGLPQRYDTQAGDKGTQLSGGQKQRVAIARAI I	1504
		
	RNPKLLLLDEATSALDAESETV VQAALDQARLGRTTIVVA	
	530 540 550 560	
5MDR2FUG.SEQ	RNPKILLLDEATSALDAESETIVQAALDKVRLGRTTIVVA	1627
3mdr2FUG.seq	RNPKLLLLDEATSALDTESEKVVQEALDQARKGRTCIVVA	1624
		
	HRLSTIQNADIIAGFSGGVVVEQGTHSQLIAIKGVYHGLV	
	570 580 590 600	
5MDR2FUG.SEQ	HRLSTIRNADIIAGFSNGKIVEQGTHSQLMEIKGVYHGLV	1747
3mdr2FUG.seq	HRLSTIQNADCI AVFQGGVVVEKGTHQQLIAKKG VYHMLV	1744
		
	TKQMGFHNVEEENTAMSELSAGEKSPVEKTVSQSSIIRRK	
	610 620 630 640	
5MDR2FUG.SEQ	TMQ-TFHNVEEENTAMSELSAGEKSPVEKTVSQSSIIRRK	1864
3mdr2FUG.seq	TKQMGYHN-----	1771
		
	STRGSSFAASEGTKEEKTED	
	650 660	
5MDR2FUG.SEQ	STRGSSFÅASEGTKEEKTEE	1924
3mdr2FUG.seq	-----D	1771

	1	2		
1		40.6	1	5MDR2FUG.SEC
2	95.9		2	3mdr2FUG.seq
	1	2		

三 利用跨膜结构预测软件（网址：<http://www.ulrec3.unil.cn/software/TMPRED-form.html>）分析 *fuguMDR1* cDNA、*fuguMDR2* cDNA 和人的 *MDR1* cDNA 的跨膜结构，三者均预测有 11 次跨膜结构。

1bp-14kb 序列的跨膜结构预测：

inside→outside		outside→inside
6- 34 (29) 1750		9- 29 (21) 2127 ++
75- 95 (21) 2221		72- 90 (19) 2320 +
144- 166 (23) 1888 +		144- 166 (23) 1752
168- 185 (18) 1908		168- 186 (19) 2153 ++
(204- 221 (18) 74)		204- 221 (18) 528 ++
246- 270 (25) 2279 ++		254- 277 (24) 1611
280- 300 (21) 2217 ++		280- 297 (18) 1722
(373- 394 (22) 268 ++)		
667- 687 (21) 1104		669- 688 (20) 1370 ++
676- 699 (24) 1099 ++		
713- 731 (19) 2787 ++		713- 731 (19) 2498
811- 836 (26) 1712		815- 837 (23) 1910 +
903- 921 (19) 1649 ++		903- 921 (19) 1375
935- 953 (19) 1480 ++		938- 955 (18) 1127

——> STRONGLY preferred model: N-terminus outside

12 strong transmembrane helices, total score : 20706

from to length score orientation

1	9	29 (21)	2127	o-i
2	75	95 (21)	2221	i-o
3	144	166 (23)	1752	o-i
4	168	185 (18)	1908	i-o

5	204	221	(18)	528	o-i
6	246	270	(25)	2279	i-o
7	280	297	(18)	1722	o-i
8	667	687	(21)	1104	i-o
9	713	731	(19)	2498	o-i
10	811	836	(26)	1712	i-o
11	903	921	(19)	1375	o-i
12	935	953	(19)	1480	i-o

——> alternative model

11 strong transmembrane helices, total score : 20281

#	from	to	length	score	orientation
1	6	34	(29)	1750	i-o
2	72	90	(19)	2320	o-i
3	144	166	(23)	1888	i-o
4	168	186	(19)	2153	o-i
5	246	270	(25)	2279	i-o
6	280	297	(18)	1722	o-i
7	667	687	(21)	1104	i-o
8	713	731	(19)	2498	o-i
9	811	836	(26)	1712	i-o
10	903	921	(19)	1375	o-i
11	935	953	(19)	1480	i-o

12kb-26kb 序列的跨膜结构预测:

Inside to outside helices : 13 found

	from	to	score	center
	7 (10)	34 (30)	1680	19
	72 (75)	93 (93)	2103	83
	150 (150)	169 (169)	1975	159
	171 (171)	188 (188)	1873	180
	249 (253)	273 (273)	2311	263
	285 (285)	303 (303)	2475	293
	372 (379)	397 (397)	207	387
	662 (668)	686 (686)	1243	677
	707 (707)	725 (725)	2477	717

800 (804)	825 (821)	1804	812
892 (892)	910 (910)	1283	900
924 (924)	942 (942)	1458	932
1018 (1018)	1036 (1036)	12	1026

Outside to inside helices : 11 found

from	to	score	center
9 (9)	34 (28)	1667	19
71 (73)	90 (90)	2262	82
147 (151)	169 (169)	1740	159
171 (171)	189 (187)	2104	179
254 (254)	271 (271)	1933	262
285 (285)	303 (303)	1945	293
662 (662)	686 (680)	1555	672
710 (710)	728 (728)	2438	720
793 (802)	826 (818)	1899	810
900 (900)	919 (919)	1077	910
924 (924)	941 (941)	1212	932

inside->outside	outside->inside
7- 34 (28) 1680	9- 34 (26) 1667
72- 93 (22) 2103	71- 90 (20) 2262 +
150- 169 (20) 1975 ++	147- 169 (23) 1740
171- 188 (18) 1873	171- 189 (19) 2104 ++
249- 273 (25) 2311 ++	254- 271 (18) 1933
285- 303 (19) 2475 ++	285- 303 (19) 1945
(372- 397 (26) 207 ++)	
662- 686 (25) 1243	662- 686 (25) 1555 ++
707- 725 (19) 2477	710- 728 (19) 2438
800- 825 (26) 1804	793- 826 (34) 1899 +
892- 910 (19) 1283 ++	900- 919 (20) 1077
924- 942 (19) 1458 ++	924- 941 (18) 1212
(1018-1036 (19) 12 ++)	

2 possible models considered, only significant TM-segments used

——> STRONGLY preferred model: N-terminus inside

11 strong transmembrane helices, total score : 20297

from to length score orientation

1 7 34 (28) 1680 i-o
 2 71 90 (20) 2262 o-i
 3 150 169 (20) 1975 i-o
 4 171 189 (19) 2104 o-i
 5 249 273 (25) 2311 i-o
 6 285 303 (19) 1945 o-i
 7 662 686 (25) 1243 i-o
 8 710 728 (19) 2438 o-i
 9 800 825 (26) 1804 i-o
 10 900 919 (20) 1077 o-i
 11 924 942 (19) 1458 i-o

——> alternative model

11 strong transmembrane helices, total score : 20217

from to length score orientation

1 9 34 (26) 1667 o-i
 2 72 93 (22) 2103 i-o
 3 147 169 (23) 1740 o-i
 4 171 188 (18) 1873 i-o
 5 254 271 (18) 1933 o-i
 6 285 303 (19) 2475 i-o
 7 662 686 (25) 1555 o-i
 8 707 725 (19) 2477 i-o
 9 793 826 (34) 1899 o-i
 10 892 910 (19) 1283 i-o
 11 924 941 (18) 1212 o-i

四 利用 predator 软件(网址: <http://www.embl-heidelberg.de/cgi/argos/predator-sorv.pl>)

初步分析 fuguMDR1 cDNA 和 fuguMDR2 cDNA 的氨基酸二级结构:

fuguMDR1 cDNA 的氨基酸 α -helix (H) 和 β -sheet (E) 结构分布

1 FRFADGWDILMVTIGVLMIAIVNGLVNPLMCIVFGEMTDSFIQEAKLSQNH 50

_____EEEEEEHHHHHHHH_____EEEEEE_____HHHHHHHHHH_____

51 NTSNPTSTMSNVLRVCCIIRHLFSIYYSILGFAVLVVAYLQMSLWTLTAA 100

```

_____EEEEHHHHHHHHHHHHHHHH_____EEEEHHHHHHHHHHHHHH
. . . . .
101 RQAKRIRELFPHGIMQQDISWYDVTETGELNTRLTEWEGIGDKAGLLIQ 150
HHHHHHHHHHHHHHHH_____HHHHHHHHHH
. . . . .
151 ASTFITSFVIGFVHGWLTLVILAI SPVLGLSAALYSKLLTSFTSKEQTA 200
HHHHEEEEEEEEEEE_____EEEE_____HHHHHHHHHHHHHH_____HHHH
. . . . .
201 YAKAGAVAAEVLSSIRTVFAFSGQRKAIKRYHKNLEDARDMGIKKGVAAN 250
HHHHHHHHHHHHHHHHHEEEEE_____HHHHHHHH_____HHHHHHHH_____EEE_____
. . . . .
251 TATGFSFLMIYLSYALAFWYGTTLVLNKEYTIGNLLTVFFVVLGAYIIG 300
_____EEEEHHHHHHHHHH_____EEEE_____EEEEEEEEEE_____EEEE
. . . . .
301 QASPNVQSFASARGAAYKVYNIIDHKPNIDSFSEGYKPEYIKGDIVFQN 350
_____HHHHHHHH_____EEEEEE_____EEEEEE
. . . . .
351 IHFSYPSRPEIKILNDSFHVNRGQTIALVGSSGCGKSTTIQLLQRFYDP 400
E_____EEEE_____EEE_____EEEEEE_____EEEE_____
. . . . .
401 QKGSIFIDGHDIRSLNIRYLREMIGVVSQEPVLFATTITENIRYGRDVT 450
_____EEE_____EEEEHHHHHHHHHHHH_____EEEEEEEEEEEE_____HHH
. . . . .
451 QEEIERATKESNAYDFIMNLPDKFETLVGDRGTQLSGGQKQRIAIARALV 500
HHHHHHHHHH_____EEEEEE_____EEEE_____HHHHHHHHHH
. . . . .
501 RNPKILLLDEATSALDAESETIVQAALDKVRLGRTTIVIAHRLSTIRNAD 550
_____EEEEHHHHHHHHHHHHHHHHHHHHHHHHHH_____EEEEEEHHHH_____E
. . . . .
551 I IAGFSNCEIVECGTHSQLMEIKGVYHGLVTMQSFQKLEDLEDSDYEPWV 600
EEEE_____EEEE_____HHHHHHHHHH_____HHHHHHHHHHHH_____HHHH
. . . . .
601 AEKSQLIESFSQSSLQRRRSTRGSL LAVSEGTKEEKEKFECDQDNIEEDE 650
HHHHHHHHHH_____HHHHHHHEEEEE_____HHHHHHHH_____
. . . . .

```


FuguMDR2cDNA 的氨基酸 α -helix (H) 和 β -sheet (E) 结构分布

```

1  FRFADRWDVLLISGTVMAMVNGTVMLMCIVFGEMTDSFIYADMAQINA 50
   _____ EEEEEEE_ EEEE_ _____ EEEE_ _____ HHHHHHHHHH_

51  SGWNSTTTILNSTLQEDMQRFAYYSVLGFVLLAAYMQVSFWTITAGRQ 100
   _____ EEEEEHHHHHHHHHHHHHHH_ HHHHHHHHHHHHHHEEEE_ ___ H

101  VKRIRSLFFHCIMQQEISWFDVNDTGELNRTLDDVYKIQEGIGDKVGLL 150
   HHHHHHHHHHHHHHEEEE_ _____ EEEEE_ _____ HHHH

151  IQAYTTFITAFIIGFTTGWKLTLVILAVSPALAI SAFFSKVLASFTSKE 200
   HHHHHHHHEEEEEEE_ _____ EEEE_ _____ HHHHHHHHHHHHHHHH_ HH

201  QTAYAKAGAVAEVLSAIRTVFAFSGQTREIERVHKNL RDAKDVGVKKAI 250
   HHHHHHHHHHHHHHHHHHHHHH_ _____ HHHHHHHHHH_ _____ HHHHH

251  SSNIAMGFTFLMIYLSYALAFWYGSTLILNFEYTI GNLLTVIFVVLIGAF 300
   HHHHHHHHHH_ HHHHHHHHHH_ _____ EEEEEEE_ EEEE_ EEEEEEE_

301  SVGQTSPNIQNFASARGAAYKVYSIIDNKPNIDSF SEDGFKPDFIKGDIE 350
   _____ HHHHHHHH_ EEEEEEE_ _____ _____ E

351  FKNIHFNYP SRPEVKILNMSLSVKSQTIALVGSSGCGKSTTIQLLQRF 400
   EEEEEEE_ _____ EEEEEEE_ EEE_ EEEEEEE_ _____ HHHHHHHHH

401  YDPEEGAVFIDGHDIRSLNIRYLREMIGVVSQEPVLFATTITENIRYGR L 450
   _____ EEEE_ EEEHHHHHHHHHHHHH_ EEEEEHHHHH_ EEEE

451  DVTQEEIERATKESNAYDFIMNLPDKFETLVGDRGTQLSGGQKQRIAIAR 500
   HHHHHHHHHHHHHH_ _____ EEEE_ EEEE_ _____ HHHHHHHHH

501  ALVRNPKILLLDEATSALDAESETIVQAALDKVRLGR TTIVVAHRLSTIR 550
   HHH_ EEEHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH_ EEEEEHHHHH_

```

```

551 NADI IAGFSNGKIVEQGTHSQLMEIKGVYHGLVMTQTFHINVEENTAMSE 600
    _EEEEEE_ EEEE_ HHHHHEEEE_ EEEE_ HHHH
    .
    .
    .
601 LSAGEKSPVEKTVSQSSIIRRKSTRGSSFAASEGTKEEKTEDEDVPDVS 650
    HHHH_ EEE_ HHHHH_
    .
    .
    .
651 FFKVLHLNIPWPYILVGLICATINGAMQPVFAILFSKIITVFADPDRDS 700
    HHHHHH_ EEEE_ EEEEE_
    .
    .
    .
701 VRRKSEFISLMFVVIGCVSFVTMFLQGYCFGKSGEILTLKRLRAFTAMM 750
    _HHHHHHHHHEEEEEHHHHHH_ HHHHHHHHHHH
    .
    .
    .
751 RQDLSWYDNPQNTV GAL TTRLAADA AQVGAAGVRLATIMQNFANLGTSI 800
    HHH_ HHHHHHHHHHHHHHHHH_ HHHHHHHHHHH_ E
    .
    .
    .
801 IIAFVYGWELTLLILAVVPLIAAAGAAEIKLLAGHAAKDKKELEKAGKIA 850
    EEEEE_ HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
    .
    .
    .
851 TEAIENVRTVVSLSREP KFECLYEENLRVPYKNSQKKAHVYGLTYSFSQA 900
    HHHHHHHHEEEEE_ HHHHH_ EEEEE_ HHH
    .
    .
    .
901 MIYFAYAACFRFGAWLIEAGRMDVEGVFLVVS AVLYGAMAVGEANTFAPN 950
    HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
    .
    .
    .
951 YAKAKMAASYLMLINKKPAIDNLSEEGTSPEKYDGNVHFEGVKFNYP SR 1000
    HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
    .
    .
    .
1001 PDVTILQGLNLKVKKGETLALVGSSGCGKSTTIQLLERFYDPREGRVSLD 1050
    _EEE_ EEEE_ EEEEE_ HHHHHHH_
    .
    .
    .
1051 GVNKQLNIHWLRSQIGIVSQEPVLFDCSLAENIAYGDNSRSVSMDEIVL 1100
    _EEEEHHHHHHHHHEEEEE_ EEEEHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
    .
    .
    .
1101 AAKAANIHSFIEGLPQRYD TQAGDKGTQLSGGQKORVAIARAIIRNPKLL 1150
    HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
    
```

1151 LLDEATSALDTESEKVVQ~~ALDQARKGRT~~CI~~VVAHRLSTIQ~~NAD~~CI~~AVFQ 1200

HHHHHHHHH_HHHHHHHHHHHH_EEEEEHHHHHHH_EEEEE_

1201 GGVVVEKGT~~HQQLIAKKG~~VYHMLVTKQMGYHND 1233

_EEEE_HHHHHHHH_____

IV cosmid 124A22 中其他序列所含基因的分析

使用 blastx 软件搜寻 26kb-40kb 中与数据库中可能存在的同源序列:

24.4kb-28kb 中存在与人 carnitine octanoyltransferase 同源的基因序列:

Sequences producing significant alignments:	(bits)	Value
sp O19094 OCTC_BOVIN PEROXISOMAL CARNITINE OCTANOYLTRANSFERASE ...	130	5e-73
sp P11466 OCTC_RAT PEROXISOMAL CARNITINE OCTANOYLTRANSFERASE (C...	109	2e-62
pir A31948 carnitine octanoyltransferase, hepatic - rat >gi 20...	108	4e-56
pir A55720 carnitine O-acetyltransferase (EC 2.3.1.7) precurs...	54	6e-10
sp P43155 CACP_HUMAN CARNITINE O-ACETYLTRANSFERASE (CARNITINE A...	54	6e-10
sp P52826 CACP_COLL1 CARNITINE O-ACETYLTRANSFERASE PRECURSOR (C...	55	7e-10
sp Q03059 CLAT_MOUSE CHOLINE O-ACETYLTRANSFERASE (CHOACTASE) (C...	50	4e-09
sp P47934 CACP_MOUSE CARNITINE O-ACETYLTRANSFERASE (CARNITINE A...	53	6e-09
sp P32738 CLAT_RAT CHOLINE O-ACETYLTRANSFERASE (CHOACTASE) (CHO...	50	6e-09
pir A48319 choline O-acetyltransferase (EC 2.3.1.6) - rat	50	6e-09
sp P28329 CLAT_HUMAN CHOLINE O-ACETYLTRANSFERASE (CHOACTASE) (C...	50	1e-08

sp|O19094|OCTC_BOVIN PEROXISOMAL CARNITINE OCTANOYLTRANSFERASE (COT) >gi|2289038 (U65745)

carnitine octanoyltransferase [Bos taurus]

Length = 612

Score = 60.1 bits (143), Expect(4) = 6e-53

Identities = 26/44 (59%), Positives = 37/44 (84%)

Query: 3397 TVRPFASEKEFKATEDIVRNFQCGVGKELHQRLLRRAETRRNWV 3266

+V+PFA+E+E+K TE IV FQ G+G++L Q+LLQRA+ RRNW+

Sbjct: 38 SVKPFANEEYKNTAIVWKFQNGIGEKLQKLLQRAKGRRWL 81

Score = 94.0 bits (230), Expect(4) = 6e-53

Identities = 39/60 (65%), Positives = 45/60 (75%)
 Query: 3180 LEQWWLDAAYLEGRSPSQLTVNFAGPAPYLEHCWPPAEGTALERASICSWHMLQYWNLIR 3001
 LE+WWL+ AYL+ R PSQL VNF GPA ++EH WPP EGT LER SI WH L YW L+R
 Sbjct: 81 LEEWWLNVAYLDVRIPSQLNVNFGGPASHIEHYWPPKEGTQLERGSISLWHNLNYWQLLR 140

Score = 65.6 bits (157), Expect(4) = 6e-53
 Identities = 26/42 (61%), Positives = 37/42 (87%)
 Query: 2900 RERLAPQKAGETPLDMDQFRMLYCTCKVPGVTKDAIRSYFKT 2775
 +E+LA +K G TPLDM+QFRML+ TCK+PG+T+D+I +YF+T
 Sbjct: 141 KEKLAVEKVGNTPLDMNQFRMLFSTCKIPGITRDSIINYFRT 182

Score = 53.1 bits (125), Expect(4) = 6e-53
 Identities = 22/38 (57%), Positives = 28/38 (72%)
 Query: 2698 TELEGRCPSHLVLCRGRIFTFDALCDGQILTPPELFR 2585
 TE EG PSHL VLCRGR+ FD + +G ++T PE+ R
 Sbjct: 182 TESEGHSPSHLAVLCRGRVVFVDMHEGYLMTAPEIQR 219

Score = 51.5 bits (121), Expect = 3e-05
 Identities = 21/34 (61%), Positives = 28/34 (81%)
 Query: 2223 LARQLSYVRQCCDGNPEGEGVSALTTEERTRWAK 2122
 + RQL+Y+++ C P+G GV+ALTTEERTRWAK
 Sbjct: 217 IQRQLTYIQKKCHSEPDGPGVAALTTEERTRWAK 250

Score = 59.7 bits (142), Expect(5) = 5e-73
 Identities = 28/47 (59%), Positives = 36/47 (76%)
 Query: 1959 QAREYLISIDPHNETILELIQSSLFTICLDETQPYSTPENYTNV*AE 1819
 +AREYLIS++P N TILE IQSSL CLD+ P+ TPE+Y+ V A+
 Sbjct: 250 KAREYLISLNPENLTILEKIQSSLVFCLEDDSPHVTPEYDYSQVSAK 296

Score = 58.6 bits (139), Expect(5) = 5e-73
 Identities = 22/36 (61%), Positives = 30/36 (83%)
 Query: 1751 TQLTRESLTGDPTIRWGDKSYNSVVYSDGTFGSNCD 1644
 +Q++ + L GDPT+RWGDKSYN + +S+G FGSNCD
 Sbjct: 291 SQVSAKILNGDPTVRWGDKSYNLIAFSNGVFCGNCD 326

Score = 130 bits (323), Expect(5) = 5e-73

Identities = 73/158 (46%), Positives = 90/158 (56%)

Query: 1560 HAPYDAMVLVTMCWYVDQRIQSTGGKWKVVTIPILGAL*IEGVRY*HCDISKGVDTVRVL 1381

HAP+DAMVLV +C+YVD+ I G+WK G + VR +

Sbjct: 327 HAPFDAMVLVKVCYYVDENILENEGRWK-----GSEKVRDI 362

Query: 1380 PPPEELVFTVDEKVRSDIGRAKKQYFESVSQFCEVELALPVWPVE*QPSVVOAQLQVVC 1201

P PEELVFTVDEKV +DI +AK QYF+ VS DLQ+V

Sbjct: 363 PVPEELVFTVDEKVLNDINQAKAQYFKQVS-----DLQLV 398

Query: 1200 YAFTAFGKAAIKQKLLHPDTFIQLAMQLAYFKLHQRYG 1087

YAFT+FGK K+K+LHPDTFIQLA+QLAY++LH R G

Sbjct: 399 YAFTSFGKLLTKEKQLHPDTFIQLALQLAYRHLHGRPG 436

Score = 81.1 bits (197), Expect(5) = 5e-73

Identities = 33/40 (82%), Positives = 36/40 (89%)

Query: 1021 RPGCCYETAMTRKFYHGRTEETMRPCTVEAVKWCTAMTDPS 902

RPGCCYETAMTR FYHGRTEET+RPCTVEAV WC +M +PS

Sbjct: 434 RPGCCYETAMTRLFYHGRTEETVRPCTVEAVNWCQSMQNP 473

Score = 31.3 bits (69), Expect(5) = 5e-73

Identities = 17/37 (45%), Positives = 23/37 (61%), Gaps = 1/37 (2%)

Query: 819 KRKAMQL-AFEKHNNLMAEAQEGRGESRPSLCLSLMGR 709

+RK M L AF KHN +M + G+G R L LSL+ +

Sbjct: 478 ERKHMMLEAFKHNKMMKDCSTGKGFDRHLLGLSLIAK 515

Score = 56.2 bits (133), Expect(2) = 1e-20

Identities = 25/31 (80%), Positives = 28/31 (89%)

Query: 662 GFDRHLLGLYLIAKEEGRPVPELFLDPLYAK 570

GFDRHLLGL LIAKEEG PVPELF DPL+++

Sbjct: 502 GFDRHLLGLSLIAKEEGLPVPELFTDPLFSR 532

Score = 67.1 bits (161), Expect(2) = 1e-20

Identities = 30/42 (71%), Positives = 35/42 (82%)

Query: 463 SGGGGNFVLSLSSLVGYTTVLGAVAPMVPHGYYFFYRIREDRW 338

SGGGGNFVLS+SLVGY V G + PMV +GYGFFY IR+DR+

Sbjct: 533 SGGGGNFVLSLSSLVGYLRVQGMVPMVHNGYGFFYHIRDDRF 574

sp|P43155|CACP_HUMAN CARNITINE O-ACETYLTRANSFERASE (CARNITINE ACETYLASE) (CAT)

Length = 624

Score = 54.3 bits (128), Expect(2) = 6e-10

Identities = 31/95 (32%), Positives = 47/95 (48%)

Query: 1383 LPPPEELVFTVDEKVRSDIGRAKKQYFESVSQFCEVELALPVPVVE*QPSVVQAQDLQV 1204

LP P++L F + +++SDI +AK Q + QDL +

Sbjct: 373 LPMPKKLRFNITPEIKSDIEKAK-----QNSLIMIQDLIT 408

Query: 1203 CYAFTAFGKAAIKQKLLHPDTFIQLAMQLAYFKLH 1099

F FGK K +KL PD FIQ+A+QLAY++++

Sbjct: 409 VMVFHFHFGKDFPKSEKLSPDAFIQMALQLAYRIY 443

Score = 32.8 bits (73), Expect(2) = 6e-10

Identities = 14/35 (40%), Positives = 24/35 (68%)

Query: 1006 YETAMTRKFYHGRTETMRPCTVEAVKWCTAMTDPS 902

YE+A R F+ GRT+T+R +++++ + AM D S

Sbjct: 450 YESASLRMFHLGRTDTIRSASMSLTFVKAMDDSS 484

28kb-32kb 中存在与人乙酰辅酶 A 结合抑制蛋白基因同源的基因序列:

sp P31786 ACBP_MOUSE	ACYL-COA-BINDING PROTEIN (ACBP) (DIAZEPAM ...	61	2e-16
sp P11030 ACBP_RAT	ACYL-COA-BINDING PROTEIN (ACBP) (DIAZEPAM BI...	60	4e-16
sp P12026 ACBP_PIG	ACYL-COA-BINDING PROTEIN (ACBP) (DIAZEPAM BI...	60	8e-16
pir S63595	acyl-coenzyme A binding protein - chicken >gi 14784...	61	1e-15
pir S63592	acyl-coenzyme A binding protein - dog >gi 1478487 b...	60	1e-15
sp P45883 ACBP_RANRI	ACYL-COA-BINDING PROTEIN HOMOLOG (ACBP) (D...	62	3e-15
sp P07108 ACBP_HUMAN	ACYL-COA-BINDING PROTEIN (ACBP) (DIAZEPAM ...	62	3e-15
pir S63594	acyl-coenzyme A binding protein - duck >gi 1478489 ...	60	3e-15
sp P45882 ACBP_ANAPL	ACYL-COA-BINDING PROTEIN (ACBP) (DIAZEPAM ...	63	4e-15
pdb 2ABD	Acyl-Coenzyme A Binding Protein (Acbp) >gi 493800 p...	59	4e-15
sp P07107 ACBP_BOVIN	ACYL-COA-BINDING PROTEIN (ACBP) (DIAZEPAM ...	59	4e-15
gi 181478	(M14200) diazepam binding inhibitor [Homo sapiens]	61	7e-15
pir S63593	acyl-coenzyme A binding protein - turtles >gi 14784...	58	2e-14
pir S48040	acyl-CoA binding protein - rape >gi 1076441 pir S4...	50	2e-09
pir A60212	endogenous anti-morphine peptide - dog	59	2e-09
gi 1006831	(U35015) acyl-CoA-binding protein [Gossypium hirsutum]	50	2e-08
gi 488839	(Z21846) diazepam binding inhibitor [Rattus norvegicus]	58	3e-07

sp|P31786|ACBP_MOUSE ACYL-COA-BINDING PROTEIN (ACBP) (DIAZEPAM BINDING INHIBITOR) (DBI)

(ENDOZEPINE) (EP) >gi|280950|pir||A60059 endozepine -
mouse >gi|50713 (X61431) diazepam-binding inhibitor [Mus
musculus]

Length = 87

Score = 60.9 bits (145), Expect(2) = 2e-16

Identities = 27/42 (64%), Positives = 36/42 (85%)

Query: 2564 QAEFEKIAEDVKKVKTRPTDQELLDLYGLYKQAIVGDVNTGR 2439

QAEF+K AE+VK++KT+PTD+E+L +Y +KQA VGDVNT R

Sbjct: 3 QAEFDKAAEEVKRLKTQPTDEMLFIYSHFKQATVGDVNTDR 44

Score = 48.4 bits (113), Expect(2) = 2e-16

Identities = 20/25 (80%), Positives = 21/25 (84%)

Query: 2359 DRPGLLDLKGKAKWDAWESRKGTIK 2285

DRPGLLDLKGKAKWD+W KGT K

Sbjct: 43 DRPGLLDLKGKAKWDSWNKLGTSK 67

sp|P07108|ACBP_HUMAN ACYL-COA-BINDING PROTEIN (ACBP) (DIAZEPAM BINDING INHIBITOR) (DBI)

(ENDOZEPINE) (EP) >gi|71950|pir||NZHU endozepine - human
>gi|181961 (M15887) endozepine precursor [Homo sapiens]

Length = 87

Score = 61.7 bits (147), Expect(2) = 3e-15

Identities = 27/42 (64%), Positives = 36/42 (85%)

Query: 2564 QAEFEKIAEDVKKVKTRPTDQELLDLYGLYKQAIVGDVNTGR 2439

QAEFEK AE+V+ +KT+P+D+E+L +YG YKQA VGD+NT R

Sbjct: 3 QAEFEKAAEEVRLKTKPSDEMLFIYGHYKQATVGDINTER 44

Score = 43.4 bits (100), Expect(2) = 3e-15

Identities = 17/25 (68%), Positives = 19/25 (76%)

Query: 2359 DRPGLLDLKGKAKWDAWESRKGTIK 2285

+RPG+LD GKAKWDAW KGT K

Sbjct: 43 ERPGMLDFTGKAKWDAWNEKGTSK 67

32.4kb-33.2k 中存在与人 Rnase P 蛋白 P38 同源的基因序列:

	Score	E
Sequences producing significant alignments:	(bits)	Value
gi 1885379 (U77664) RNaseP protein P38 [Homo sapiens]	155	3e-37
gi 1885379 (U77664) RNaseP protein P38 [Homo sapiens]		
Length = 283		
Score = 155 bits (387), Expect = 3e-37		
Identities = 97/237 (40%), Positives = 133/237 (55%), Gaps = 18/237 (7%)		
Query: 18 RYIPTKTCFTSPFTPKWSPLPQEDMHFILNLTLENFVSIGLVKKEPKVFR—PWRKKKK 188		
R + KT +P+ +WS L EDMHFIL TL++ +IGL K E K + P+ KK+		
Sbjct: 17 RPLVVKTSLNPNYIIRWSALESEDMHFILQTLERLKAIGLQKIEDKKKNKTPFLKES 76		
Query: 189 QEX-XXXXXXXXXXXXXXXXXEPKRGWTDVAARRKLAIGINEVTKALERNEXXXXXXXX 365		
+E + GWT R++L IG+NEVT+ALER E		
Sbjct: 77 REKCSIAVDISDNLKEKKTDAKQVSGWTPAHVRKQLVIGVNEVTRALERRELLVLVCK 136		
Query: 366 XXXPQHMEHLITLSTTRDVPACQVPRLSQSVSEPLGLKSVLALGFRQCLPQERDVFSNV 545		
P + HLI LS +R VPACQVPRLS+ ++ +GLK VLAL F+ + F+		
Sbjct: 137 SVKPAMITSHLIQLSLSRSPACQVPRLSERAPVIGLKCVLALAFK—KNTDFVDE 192		
Query: 546 VEAILPKVPPLDVPWLQ—DTPASIKPDENRQKRRLETESEEGTP—VS 683		
V AI+P+VP L VPWLQ D+ +++ + Q R L S E +		
Sbjct: 193 VRAIIPRVPSLSVPWLQDRIEDSGENLETEPLESQDRELLDTSFEDLSKPKRKLADGRQA 252		
Query: 684 STTLQPLKVKKIVPN 728		
S TLQPLK+KK++PN		
Sbjct: 253 SVTLQPLKIKKLIPN 267		

36kb-40kb 中存在与人 N-肉豆蔻酰转移酶 2 同源的基因序列:

	Score	E
Sequences producing significant alignments:	(bits)	Value
gi 3005065 (AF043325) N-myristoyltransferase 2 [Homo sapiens]	127	1e-80
gi 3005069 (AF043327) N-myristoyltransferase 2 [Mus musculus]	127	9e-80
gi 3005067 (AF043326) N-myristoyltransferase 1 [Mus musculus]	125	4e-79
gi 2443814 (AF020500) myristoyl CoA:protein N-myristoyltransfer...	125	8e-79
sp P30419 NMT_HUMAN GLYCYLPEPTIDE N-TETRADECANOYLTRANSFERASE (P...	125	8e-79
gi 3005063 (AF043324) N-myristoyltransferase 1 [Homo sapiens]	125	8e-79

gi|2996202 (AF053725) myristoyl-CoA: protein N-myristoyl transf... 101 7e-60
 sp|P46548|NMT_CAEEL PROBABLE GLYCYLPEPTIDE N-TETRADECANOYLTRANS... 98 8e-46
 sp|P34809|NMT_CRYNE GLYCYLPEPTIDE N-TETRADECANOYLTRANSFERASE (P... 92 4e-40
 sp|P30418|NMT_CANAL GLYCYLPEPTIDE N-TETRADECANOYLTRANSFERASE (P... 87 5e-35
 gn1|PID|e1263902 (AL022103) hypothetical protein [Schizosacchar... 86 7e-35
 sp|P34763|NMT_AJECA GLYCYLPEPTIDE N-TETRADECANOYLTRANSFERASE (P... 88 5e-33
 sp|P14743|NMT_YEAST GLYCYLPEPTIDE N-TETRADECANOYLTRANSFERASE (P... 87 2e-31
 gi|806885 (U25270) myristoyl-CoA:protein N-myristoyltransferase... 87 2e-31
 gi|972247 (F14725) glycylopeptide N-tetradecanoyltransferase [Su... 100 7e-20
 pir||S30363 myristoyl-CoA—protein N-myristoyltransferase - bovine 51 5e-05

gi|3005065 (AF043325) N-myristoyltransferase 2 [Homo sapiens]

Length = 498

Score = 50.8 bits (119), Expect = 5e-05

Identities = 23/37 (62%), Positives = 24/37 (64%)

Query: 3582 MAEDXXXXXXXXXXLDDQDTGIDGDNEEENEHLGG 3472

MAED LDDQDTGIDGDNEEE EH +G

Sbjct: 1 MAEDSESAASQSLLEDDQDTGIDGDNEEETEHAHG 37

Score = 90.9 bits (222), Expect = 4e-17

Identities = 41/51 (80%), Positives = 49/51 (95%), Gaps = 1/51 (1%)

Query: 2320 NPTLPIQKLDIQRAMELLS-CQGPASIDEAAKHKHYQFWDTQPVPKLSKCL 2168

NP++P+QKLDIQRAMELLS CQGPA++IDEAAKH+YQFWDTQPVPKL + +

Sbjct: 83 NPSVPMQKLDIQRAMELLSACQGPARNIDEAAKHRYQFWDTQPVPKLDEVI 134

Score = 80.0 bits (194), Expect(3) = 3e-73

Identities = 34/45 (75%), Positives = 40/45 (88%)

Query: 1867 DEVVTSHGPIEADKENIRQEPYSLPQGFMWDTLDLGSAEVVRRLW 1733

DEV+TSHG IE DK+N+RQEPYSLPQGFMWDTLDL AEV++ L+

Sbjct: 131 DEVITSHGAIPEKDNVRQEPYSLPQGFMWDTLDLSDAEVLKELY 175

Score = 127 bits (317), Expect(3) = 3e-73

Identities = 65/98 (66%), Positives = 67/98 (68%)

Query: 1641 LKELYTLLNENYVEDDDNMFREFDYSPNFKW*APCSAFLPGSLTARRPQGAELVSPRA 1462

LKELYTLLNENYVEDDDNMFREFDYSP FL W A

Sbjct: 171 LKELYTLLNENYVEDDDNMFREFDYSPFELLW-----A 202

Query: 1461 LRPPGWLQWHCGVRVSSNKKLVGFISAIPADIRIYDT 1348

LRPPGWL QWHCGVRVSSNKKLVGFISAIPA+IRIYD+

Sbjct: 203 LRPPGWLLQWHCGVRVSSNKLVGFISAIPANIRIYDS 240

Score = 113 bits (279), Expect(3) = 3e-73

Identities = 54/57 (94%), Positives = 57/57 (99%)

Query: 1256 VKRMVEINFLCVHKKLRSKRVAPVLIREITRRVNLEGIFQAVYTAGVVLPKPVSTCR 1086

VK+MVEINFLCVHKKLRSKRVAPVLIREITRRVNLEGIFQAVYTAGVVLPKP++TCR

Sbjct: 241 VKKMVEINFLCVHKKLRSKRVAPVLIREITRRVNLEGIFQAVYTAGVVLPKPIATCR 297

Score = 89.3 bits (218), Expect(4) = 1e-80

Identities = 40/43 (93%), Positives = 42/43 (97%)

Query: 897 CRYWHRSLNPRKLVEVKFSLSRNMTLQRTMKLYRLPDVSSSTN 769

CRYWHRSLNPRKLVEVKFSLSRNMTLQRTMKLYRLPDV+ T+

Sbjct: 296 CRYWHRSLNPRKLVEVKFSLSRNMTLQRTMKLYRLPDVTKTS 338

Score = 83.9 bits (204), Expect(4) = 1e-80

Identities = 39/64 (60%), Positives = 49/64 (75%)

Query: 697 TKTPGLRPMERRDIRQVTELLQKFLKRFQLAPSMTEEEVSHWFLPDNIIDTYVVEVAGI 518

TKT GLRPME +DI+ V EL+ +LK+F LAP M EEEV+HWFLP+++IIDT+VVE

Sbjct: 335 TKTSGLRPMEPKDIKSVRELINTYLKQFHLAPVMDEEEVAHWFLPREHIIDTFVVEPNG 394

Query: 517 SLKD 506

L D

Sbjct: 395 KLTD 398

Score = 99.8 bits (245), Expect(4) = 1e-80

Identities = 45/56 (80%), Positives = 50/56 (88%)

Query: 446 QGAGGALTDFAFYTLPSTVMHHPHRSKAAYSFYNVHTQTPLLDLMDALILAK 279

+ G LTDF SFYTLPSTVMHHP H+SLKAAYSFYN+HT+TPLLDLM+DALILAK

Sbjct: 390 ESPNGKLTDFLSFYTLPSTVMHHPAHKSLKAAYSFYNIHTETPLLDLMSDALILAK 445

Score = 93.6 bits (229), Expect(4) = 1e-80

Identities = 40/49 (81%), Positives = 44/49 (89%)

Query: 205 RQKGFDFVFNALDLMENKVFLEKLFKFGIGDGNLQYYLYNWKCPSMEPKV 59

+ KGFDFVFNALDLMENK FLEKLFKFGIGDGNLQYYLYNW+CP + +KV

Sbjct: 445 KSKGFDFVFNALDLMENKTFLEKLFKFGIGDGNLQYYLYNWRCPGTDSEKV 493

综上所述: 在 15.5kb 的河豚鱼基因组序列中存在着 4 个与人同源的基因序列, 另外, 在 1bp-1kb 中存在着一段与人 MDRI cDNA 3 端同源的序列, 该序列的终止密码子在 489bp-491bp, 为 TAA, 在 980bp-985bp 处有加尾信号 AATAAA。

讨 论

人类基因组计划的前期计划是获得四张图：遗传图、物理图、转录图和序列图。目前该计划正如期甚至提前进展。到 2005 年获得人类基因组 3×10^9 bp 的任务将会圆满完成，但这并不是人类基因组计划的最终目标。 3×10^9 个由 A、G、C、T 组成的序列图并不能告诉我们什么是基因，什么是基因组，更不能告诉我们什么是调控、表达序列，也就是说，我们得到了决定人生老病死的一本天书，但并没有破译天书的钥匙。人类基因之间或者说真核生物基因虽然千变万化，但都有一些共同的特性如剪接位点一般遵循“ag.....gt”法则，起始密码子永远是 ATG，终止密码子是 UAA、UAG 或 UGA。启动子序列区已有一些明显的特征如 TATA box 等结构。这些特征的捕获辅之以计算机技术，产生了生物信息学这门新兴的边缘学科 (Bioinformatics)，这就是打开这本天书的最初钥匙。

研究一个基因的结构功能、表达调控，最好是具有该基因的全长序列和它的全长 cDNA 序列。但一般情况下，我们较难获得全长 cDNA 序列。借助于生物信息技术，我们可以对一段未知 DNA 序列进行结构模拟分析：获取初步资料，即预测 cDNA，再以该 cDNA 为模板对其二级结构、三级结构进行模拟，最终预测该未知基因的功能。对于 cDNA 序列已知、甚至功能已知的基因，我们也可以借助于生物信息技术，分析基因和蛋白序列的关键作用部位，可以模拟显示各种突变对基因功能的影响。从而为基因功能的研究开辟了一条“捷径”。

本文所涉及的基因—多药耐药基因是一个功能已知的基因。该基因广泛存在于各种生物中，从原核生物到真核生物，其功能详见综述。但对于河豚鱼中是否存在该基因，该基因在河豚鱼中是否也起着同样作用，我们尚无法得知。本文最初获得的仅是有与人同源的河豚鱼多药耐药基因 cosmid 的全长序列，在这 40kb 序列中，其可能存在的多药耐药基因位于何处不得而知。生物信息学为该问题的解决提供了较大的方便。

首先我们先通过 Dotplot 软件系统粗略分析，找到 contig2 中可能存在与 MDR1 cDNA 同源的基因序列位于何处，以缩小下一步探索范围。在获得两段序列范围之后，通过 blastx 可较直接方便的获知这两段序列中何处的翻译序列与 p-glycoprotein 有同源性。目前，国际上一般流行使用 Grail 软件系统来预测一段未知序列中可能存在的编码序列即外显子序列。本文尝试使用该软件系统时，发现对于 1bp—14kb 这段序列，使用 organism=human 时的 Grail1 程序得到的翻译序列与人 p-glycoprotein 有着较高的同源性；但对于 12kb—26kb 序列，只有使用

organism=mouse 时的 *graila* 程序分析得到的翻译序列与人 *p-glycoprotein* 有着较多的同源性。由于河豚鱼作为最原始的脊椎动物，自然与人或鼠有着很大的进化距离，利用人或鼠的剪接特点套用于河豚鱼，必然会造成一些误差，加上本身河豚鱼的两个多药耐药基因是否是同样的起源。就象人的 *MDR1* 和 *MDR3* 一样，现在还不得而知。这可能用来解释上述为什么 *grail* 分析的结果与人的同源性并不是很高，且两种多药耐药基因需采取不同的生物模型才能获得满意的结果。

1bp-14kb 的 *Grail* 分析结果

[*graillexons* -> Exons]

	Str	Frm	Start	End	ORFstart	ORFend	Score	Quality
1-	f	2	101	221	4	233	9.299	excellent
2-	f	2	291	481	234	491	15.361	excellent
3-	f	2	2613	2761	2616	2963	12.580	excellent
4-	f	0	3521	3641	3397	4065	12.023	excellent
5-	f	1	4161	4221	4133	4261	2.947	good
6-	f	0	4411	4541	4357	4584	10.446	excellent
7-	f	1	4561	4763	4553	4765	20.359	excellent
8-	f	0	5611	5691	5584	5811	7.006	excellent
9-	f	2	5801	5911	5718	5978	6.239	excellent
10-	f	2	6261	6431	6255	6890	16.126	excellent
11-	f	2	6541	6761	6255	6890	17.942	excellent
12-	f	1	7111	7231	7073	7399	11.060	excellent
13-	f	2	7351	7401	7245	7607	4.795	good
14-	f	0	7801	7901	7675	7971	7.064	excellent
15-	f	1	9591	9671	9539	9697	4.967	good
16-	f	2	9741	9921	9702	9977	16.474	excellent
17-	f	1	10041	10091	9983	10135	1.737	marginal
18-	f	2	11261	11351	11217	11444	6.767	excellent
19-	f	2	11491	11641	11460	11714	10.692	excellent
20-	f	0	12121	12601	12073	12612	26.477	excellent
21-	f	0	12671	12868	12613	12870	18.584	excellent
22-	r	2	13321	13341	13299	13367	1.170	marginal
23-	r	1	10591	10651	10553	10663	3.543	good
24-	r	2	10151	10321	10098	10454	15.397	excellent
25-	r	0	6671	6701	6664	6834	1.340	marginal

26-	r	1	5534	5611	5537	5647	2.529	good
27-	r	2	1911	1981	1914	2009	5.713	excellent

[graillexons -> Exon Translations]

28- YDTQAGDKGTQLSGGQKQKRVAIARAIIRNPKLLLLDEATSALDTESEKVR
LYSH

29- QHCTSCVQVQVEALDQARKGRTCIVVAHRLSTIQNADCIAVFQGGVVVEK
GTHQQLIAKKGVYHMLVTKQMGQHSQ

30- FRFADGWDILMVTIGVLMIAVNGLVNPLMCIVFGEMTDSFIQEAKLSQNH
NTSNPSEC

31- SILGFAVLVVAYLQMSLWTLTAARQAKRIRELFFHGIMQQDISWYDVTET
GELNTRLTEW

32- LLTSFTSKEQTAYAKAGAVAAEVLSSIRTVFAFSGQRKA

33- HKNLEDARDMGIKKGAANTATGFSFLMIYLSYALAFWYGTTLVLNKEYT
IGNLLTVSGGRHH

34- VEADIITSTIKNKSVAAEVTTTCVQMKVFFVLYGAYIIGQASPNVQSFA
SARGAAYKVYNIIDHVSPV

35- SISAISQKPNIDSFSEGGYKPEYIKGDIVFQNIHFSYPSRPEIK

36- HVPAPTHLSHPQILNDMSFHRNGQTIALVGSSCGKSTTIQLLQRFYDP
QKGSVSN

37- IFIDGHDIRSLNIRYLREMIQVVSQEPVLFATTITENIRYGRLDVDTQEEI
ERATKESNAYDFIMNLPD

38- NANIYIMISEQKFETLVGDRGTQLSGGQKQRIAIARALVRNPKILLDEA
TSALDAESETIVQAALDKVQTQTRAEMMGNKETLLHCLLVVYS

39- SLQVRLGRTTIVIAHRLSTIRNADIIAGFSNGEIVEGTHSQLMEIKGVY
HGLVTMQVS

40- LGHLFLLQSFQKLEDLEDSDYEPWVAEKSLIESFSQ

41- VSFFKVMRYNVSEWPYILVGTICAMINGAMQPVFSIIFTEIIMVHYVPVF
TFL

42- LKDSLWYDNPKNVTGALTTRLAADAAHVQGVSLIRDVTFDLSEA

43- PLSAQAAGVRLAVMTQNFANLGTSSIIISFVYGWELTLLILAVVPILAVAG
AAEVKLLTGAAEDKKELEMAGKVFLFQFL

44- RGFSLQIATEAIENVRTVVSLTREPTFVALYEENLT

45- TLFPSVVMTMLYGAMAVGEANTYAPNFAKAKISASHLTMLINRQPAIDNL

46- HTVLLCRLQEKYDGNVLFEDVKPNYSPRPDVPVLQGLNLEVQKGETLALV

GSSGCGKSTTIQLLERFYDP

47- LVPLHILYITFVLSQLLDGVDVKQLNVHWLRSQIGIVSQEPVLFDCSLAE
 NIAYGDNSRSVSMDEIVAAKAANIHSFIEGLPQVAAVNQGKWLIPHLID
 SHGAAHDHLHHIQTVSEQRDYDQAGDKGTQLSGGQKQRVATARAII RNPK
 LLLLDEATSALDTESEKVRLYSH

48- QHCTSCVQVQVEALDQARKGRTCIVVAHRLSTIQNADCIAVFQGGVVVEK
 GTHQQLIAKKGVYHMLVTKQMGYHSG

[trans_start= 4381]

HKNLEDARDMGIKKGVAANTATGFSFLMIYLSYALAFWYGTTLVLNKEYT
 IGNLLTVSGRHH

49- VEADIITSTIKNKSVAETVTCVQMKVFFVVLGAYIIGQASPNVQSFA
 SARGAAYKVYNIIDHVSPV

50- SISAISQKPNIDSFSEGGYKPEYIKGDIVFQNIHFSYPSRPEIK

51- HVPAPTHLSHPQILNDMSFHVRNGQTIALVGSSGCGKSTTIQLLQRFYDP
 QKGSVSN

52- IFIDGHDIRSLNIRYLREMIQVVSQEPVLFATTITENIRYGRDLVTQEEI
 ERATKESNAYDFIMNLPD

53- NANIYIMISEQKFETLVGDRGTQLSGG[End translation]

[Start translation]

[trans_start= 13299]

TGPELTEESSELKRKIDHDNEL

54- TRLGKEATSLYRLAIKMSERTGADMVGDLLRGW

55- GQDDQGELPSMYKTNNKGSDDKGGRLDQQTRLVTDALLDLVDIVRTRESN
 TTYDNDHASC DRNWSVYDVRHPLCETR

56- QEKMSQDGRSDYTQSTTETHVD

57- HLWDLLERETHIMDVQLMLMLSCQNGPKSPRNIS

58- TLHFALNINMDLVLEKADDHVPSLIYCGEG

Grail 结果的 blastp 分析结果

sp|P08183|MDR1_HUMAN MULTIDRUG RESISTANCE PROTEIN 1 (P-GLYCOPROTEIN 1)

Length = 1280

Score = 1051 bits (2689), Expect = 0.0

Identities = 640/1432 (44%), Positives = 813/1432 (56%), Gaps = 396/1432 (27%)

Query: 129 SGFRFADGWDILMVTIGVLMIAVNLVNLPLMCIVFGEMTDSFIQEAKLSQ-NHNTSNPSE 187

S FR+++ D L + +G L AI++G PLM +VFGEMTD F L N +N S+

Sbjct: 38 SMFRYSNWLKLYMVVGTAAIIHGAGLPLMMLVFGEMTDIFANAGNLEDLMSNITNRS 97

Query: 188 CSILGF-----AVLVVAYLQMSLWTLTAARQAKRIRELFHFGIMQ 227
 + GF VLV AY+Q+S W L A RQ +IR+ FFH IM+

Sbjct: 98 INDTGFFMNL EEDMTRYAYYYSGIGAGVLVAAYIQVSFWCLAAGRQIHKIRKQFFHAIMR 157

Query: 228 QDISWY-----DVTETGE----- 240
 Q+I W+ DV++ E

Sbjct: 158 QEIGWFDVHDV GELNTRLTDDVSKINEGIGDKIGMFFQSMATFFTFGIVGFTRGWKLT LV 217

Query: 241 -----LNRLTEWLLTSFTSKEQTXXXXXXXXXXEVLSSIRTVFAFSGQRKA----- 287
 L+ + +L+SFT KE EVL++IRTV AF GQ+K

Sbjct: 218 ILAISPVLGLSAAVWAKILSSFTDKELLAYAKAGAVAEVLA AIRTVIAFGGQKKE LERY 277

Query: 288 HKNLEDARDMGIKKGVAANTATGFSFLMIYLSYALAFWYGTTLVLNKEYTIGNLLTVSGG 347
 +KNLE+A+ +GIKK + AN + G +FL+IY SYALAFWYGTTLV L+ EY+IG +LTV

Sbjct: 278 NKNLEEAKRIGIKKAITANISIGAAFLLIYASYALAFWYGTTLVLSGEYSIGQVLT V----- 334

Query: 348 RHHVEADIITSTIKNKSVA AETVTTCVQMKVFFVLYGAYIIGQASPNVQSFASARGAAY 407
 FF VL GA+ +GQASP++++FA+ARGAAY

Sbjct: 335 -----FFSVLIGAFSVGQASPSIEAFANARGAAY 363

Query: 408 KVYNIIDHVSPVVSISAISQQKPNIDSFSEGDYKPEYIKGDIVFQNIHFSYPSRPEIKHV 467
 +++ IID+ KP+IDS+S+ G+KP+ IKG++ F+N+HFSYPSR E+K

Sbjct: 364 EIFKIIDN-----KPSIDSYSKSGHKPDNIKGNLEFRNVHFSYPSRKEVK-- 408

Query: 468 PAPTHLSHPQILNDMSFHVRNGQTIALVGS GCGKSTTIQLLQRFYDPQKGSVSNIFIDG 527
 IL ++ V++GQT+ALVG+SGCGKSTT+QL+QR YDP +G VS +DG

Sbjct: 409 -----ILKGLNLKVQSGQTVALVGN S GCGKSTTVQLMQRLYDPT EGMVS---VDG 455

Query: 528 HDIRSLNIRYLREMI GVV SQEPVLFATTITENIRYGR LDVTQEEIERATKESNAYDFIMN 587
 DIR++N+R+LRE+IGVVSQEPVLFATTI ENIRYGR +VT +EIE+A KE+NAYDFIM

Sbjct: 456 QDIRTINVRFLREIIGVVSQEPVLFATTIAENIRYGRENVTMDEIEKAVKEANAYDFIMK 515

Query: 588 LPDNANIYIMISEQKFETLVGDRGTQLSGGQKQRIAIARALVRNPKILLLDEATSALDAE 647
 LP KF+TLVG+RG QLSGGQKQRIAIARALVRNPKILLLDEATSALD E

Sbjct: 516 LP-----HKFDTLVGERGAQLSGGQKQRIAIARALVRNPKILLLDEATSALDTE 564

Query: 648 SETIVQAALDKVQTQTRAE MGNKETLLHCLLVYVSSLQVRLGRTTIVIAHRLSTIRNAD 707
 SE +VQ ALDK R GRTTIVIAHRLST+RNAD

Sbjct: 565 SEAVVQVALDKA-----RKGRTTIVIAHRLSTVRNAD 596

Query: 708 IIAGFSNGEIVEQGT HSQLMEIKGVYHGLVTMQVSLGHLFLL-----QSFQKLEDLSD- 762
 +IAGF +G IVE+G H +LM+ KG+Y LVTMQ + + L +S +++ LE S

Sbjct: 597 VIAGFDDGVIVEKGNHDELMKEKGIYFKLVMTQTAGNEVELENAADESKSEIDALEMSSN 656

Query: 763 -----YEPWVAEKSQLIESFSQVSFFKVMRYNVSEWPYILVGTI 801
 + ++ K L ES VSF+++M+ N++EWPY +VG

Sbjct: 657 DSRSSLIRKRSTRRSVRCGSAQDRKLSLKEALDESTIPPVSFWRIMKLNLTWPYFVGVF 716
 Query: 802 CAMINGAMQPVFSIIFTEIIMV----- 823
 CA+ING +QP F+IIF++II V

Sbjct: 717 CATINGGLQPAFAIIFSFIIGVFRIDDPETKRQNSNLSLLFLALGIISFITFFLQGF 776
 Query: 824 -----HYPVPFTFLKDLSDWYDNPKNVTGALTTRLAADAHVQGVSLIRDVTF 871
 Y V + L +D+SW+D+PKNT GALTRLA DAA V+G

Sbjct: 777 FKGAGEILTKRLRYMVFRSMLRQDVSWFDDPKNTTGALTTRLANDAAQVKG----- 827
 Query: 872 DLSEAPLSAQAGVRLAVMTQNFANLGTSTIIISFVYGWXXXXXXXXXXXXXXXXXXEVK 931
 A G RLAV+TQN ANLGT IIISF+YGW+ E+K

Sbjct: 828 -----AIGSRLAVITQNIANLGTGIIISFIYGWQLTLLLLAIVPIIAIAGVEMK 877
 Query: 932 LLTGHAEDKKELEMAGKVFLFQFLRGFSLQIATEAIENVRTVVSLTREPTFVALYEENL 991
 +L+G A +DKKELE AGK IATEAIEN RTVVSLT+E F +Y ++L

Sbjct: 878 MLSGQALKDKKELEGAGK-----IATEAIENFRTVVSLTQEQKFEMYAQSL 924
 Query: 992 -----TTLFPSVVM-----TMYG 1005
 T F +M +++G

Sbjct: 925 QVPYRNSLRKAHIFGITFSFTQAMMYFSYAGCFRFGAYLVAHKLMSFEDVLLVFSAVVFG 984
 Query: 1006 AMAVGEANTYAPNFAKAKISASHLTMLINRQPAIDNLHTVLLCRLQEKYDGNVLFEDVKF 1065
 AMAVG+ +++AP++AKAKISA+H+ M+I + P ID+ T L + +GNV F +V F

Sbjct: 985 AMAVGQVSSFAPDYAKAKISAAHIIMIIEKTPLIDSYSTEGL--MPNTLEGNTFGEVVF 1042
 Query: 1066 NYSPRPDPVVLQGLNLEVQKGETLALVGSSGCGKSTTIQLLERFYDPLVPLHILYITFVL 1125
 NYP+RPD+PVLQGL+LEV+KG+TLALVGSSGCGKST +QLLERFYDPL

Sbjct: 1043 NYPTRPDIPVLQGLSLEVKKGQTLALVGSSGCGKSTVVQLLERFYDPLAG----- 1092
 Query: 1126 SQLLDGVDVKQLNVHWLRSQIGIVSQEPVLFDCSLAENIAYGDNSRVSMDIIVAAKAA 1185
 LLDG ++K+LNV WLR+ +GIVSQEP+LDCS+AENIAYGDNSR VS +EIV AAK A

Sbjct: 1093 KVLLDGKEIKRLNVQWLRHLGIVSQEPILFDCSLAENIAYGDNSRVVSQEEIVRAAKEA 1152
 Query: 1186 NIHSFIEGLPQVAAVNQGKWLIPHLIDSHGAAHDHLHHIQTVSEQRDYDTQAGDKGTQLSG 1245
 NIH+FIE LP +Y T+ GDKGTQLSG

Sbjct: 1153 NIHFIESLP-----NKYSTKVGDKGTQLSG 1178
 Query: 1246 GQKQRAIARAIIIRNPKLLLLDEATSALDTESEKVRLYSHQHCTSCVQVQVQALDQARKG 1305
 GQKQR+AIARA++R P +LLLDEATSALDTESEK VVQVQALD+AR+G

Sbjct: 1179 GQKQRIATARALVRQPHIILLDEATSALDTESEK-----VVQVQALDKAREG 1224
 Query: 1306 RTCIVVAHRLSTIQNADCIAVFQGGVVVEKGTHTQQLIAKKGVYHMLVTKQMG 1357
 RTCIV+AHRLSTIQNAD I VFQ G V E GTHQQL+A+KG+Y +V+ Q G

Sbjct: 1225 RTCIVIAHRLSTIQNADLIVVFQNGRVKEHGTHQQLLAQKGIYFSMVSQAG 1276

Score = 268 bits (677), Expect = 7e-71

Identities = 136/244 (55%), Positives = 180/244 (73%), Gaps = 39/244 (15%)

Query: 1387 MKVFFVVLGAYIIGQASPNVQSFASARGAAYKVYNIIDHVSPVVSISAISSQKPNIDSF 1446
 + VFF VL GA+ +GQASP++++FA+ARGAAY+++ IID+ KP+IDS+

Sbjct: 332 LTVFFSVLIGAFSVGQASPSIEAFANARGAAYEIFKIIDN-----KPSIDSY 378

Query: 1447 SEDGYKPEYIKGDIVFQNIHFSYPSRPEIKHVPAPTHLSHPQILNDMSFHVRNGQTIALV 1506
 S+ G+KP+ IKG++ F+N+HFSYPSR E+K IL ++ V++GQT+ALV

Sbjct: 379 SKSGHKPDNIKGNLEFRNVHFSYPSRKEVK-----ILKGLNLKVQSGQTVALV 426

Query: 1507 GSSGCGKSTTIQLLQRFYDPQKGSVSNIFIDGHDIRSLNIRYLREMIGVVSQEPVLFATT 1566
 G+SGCGKSTT+QL+QR YDP +G VS +DG DIR++N+R+LRE+IGVVSQEPVLFATT

Sbjct: 427 GNSGCGKSTTVQLMQRLYDPTEGMVS---VDGDIRTINVRFLREIIGVVSQEPVLFATT 483

Query: 1567 ITENIRYGRDVTQEEIERATKESNAYDFIMNLPDNANIYIMISEQKFETLVGDRGTQLS 1626
 I ENIRYGR +VT +EIE+A KE+NAYDFIM LP KF+TLVG+RG QLS

Sbjct: 484 IAENIRYGRENVTMDEIEKAVKEANAYDFIMKLP-----HKFDTLVGERGAQLS 532

Query: 1627 GGTR 1630
 GG +

Sbjct: 533 GGQK 536

Score = 181 bits (455), Expect = 8e-45

Identities = 103/244 (42%), Positives = 146/244 (59%), Gaps = 41/244 (16%)

Query: 1389 VFFVVLGAYIIGQASPNVQSFASARGAAYKVYNIIDHVSPVVSISAISSQKPNIDSFSE 1448
 VF V++GA +GQ S +A A+ +A + II+ + P IDS+S

Sbjct: 977 VFSAVVFGAMAVGQVSSFAPDYAKAKISAAHIIMIIE-----KTPLIDSYST 1023

Query: 1449 DGYKPEYIKGDIVFQNIHFSYPSRPEIKHVPAPTHLSHPQILNDMSFHVRNGQTIALVGS 1508
 +G P ++G++ F + F+YP+RP+I +L +S V+ GQT+ALVGS

Sbjct: 1024 EGLMPNTLEGNVTFGEVVFNYPTRPDIP-----VLQGLSLEVKKGQTLALVGS 1071

Query: 1509 SGCGKSTTIQLLQRFYDPQKGSVSNIFIDGHDIRSLNIRYLREMIGVVSQEPVLFATTIT 1568
 SGCGKST +QLL+RFYDP G V +DG +I+ LN+++LR +G+VSQEP+LF +I

Sbjct: 1072 SGCGKSTVVQLLERFYDPLAGKV---LLDGKEIKRLNVQWLRHLGIVSQEPILFDCSIA 1128

Query: 1569 ENIRYGRD---VTQEEIERATKESNAYDFIMNLPDNANIYIMISEQKFETLVGDRGTQLS 1626
 ENI YG V+QEEI RA KE+N + FI +LP+ K+ T VGD+GTQLS

Sbjct: 1129 ENIAYGDNSRVVSQEEIVRAAKEANIHFIESLPN-----KYSTKVGDKGTQLS 1177

Query: 1627 GGTR 1630
 GG +

Sbjct: 1178 GGQK 1181

Score = 164 bits (412), Expect = 9e-40

Identities = 85/126 (67%), Positives = 98/126 (77%), Gaps = 14/126 (11%)

Query: 1 YDTQAGDKGTQLSGGQKQRVAIARAIIRNPKLLLLDEATSALDTESEKVRLYSHQHCTSC 60
 Y T+ GDKGTQLSGGQKQ+AIARA++R P +LLLDEATSALDTESEK
 Sbjct: 1165 YSTKVGDKGTQLSGGQKQRIAIARALVRQPHILLDEATSALDTESEK----- 1213
 Query: 61 VQVQVEALDQARKGRTCIVVAHRLSTIQNADCIAVFQGGVVVEKGTHTQQLIAKKGVYHML 120
 VQEALD+AR+GRTCIV+AHRLSTIQNAD I VFQ G V E GTHQQL+A+KG+Y +
 Sbjct: 1214 ---VQEALDKAREGRTCIVIAHRLSTIQNADLIVVFQNGRVKEHGTHQQLLAQKGIYFSM 1270
 Query: 121 VTKQMG 126
 V+ Q G
 Sbjct: 1271 VSVQAG 1276

Score = 299 bits (758), Expect = 2e-80

Identities = 204/588 (34%), Positives = 291/588 (48%), Gaps = 145/588 (24%)

Query: 834 KDLSWYDNPKNVTGALTTRLAADAAHVQGVSLIRDVTFDLSEAPLSAQAGVRLAVMTQN 893
 +++ W+D + VG L TRL D + + + G ++ + Q+
 Sbjct: 158 QEIGWFD---VHDVGEINTRLTDDVSKIN-----EGIGDKIGMFFQS 196
 Query: 894 FANLGTSTIIISFVYGWEXXXXXXXXXXXXXXXXXXEVKLLTGHAEDKKELEMAGKVFLF 953
 A T I+ F GW+ K+L+ ++ AG V
 Sbjct: 197 MATFFTGFIVGFTRGWKLTVLILAIISPVLGLSAAVWAKILSSFTDKELLAYAKAGAV--- 253
 Query: 954 QFLRGFSLQIATEAIENVRTVVSLTREPTFVALYEENL----- 991
 A E + +RTV++ + + Y +NL
 Sbjct: 254 -----AEEVLAAIRTVIAFGGQKKELERYNKNLEEAKRIGIKKAITANISIGAAF 303
 Query: 992 -----TTLFPS-----VMTMLYGAMAVGEANTYAPNFAKAKISAS 1027
 TTL S V ++L GA +VG+A+ FA A+ +A
 Sbjct: 304 LLIYASYALAFWYGTTLVLSGEYSIGQLTVFFSVLIGAFSVGQASPSIEAFANARGAAY 363
 Query: 1028 HLTMLINRQPAIDNLHTVLLCRLQEKYDGNVLFEDVKFNYPSPRPDPVVLQGLNLEVQKGE 1087
 + +I+ +P+ID+ + GN+ F +V F+YPSR +V +L+GLNL+VQ G+
 Sbjct: 364 EIFKIIDNKPSIDSYSKS---GHKPDNIKGNLEFRNVHFSYPSRKEVKILKGLNLKVQSGQ 421
 Query: 1088 TLALVGSSGCGKSTTIQLLERFYDPLVPLHILYITFVLSQLLDGVDVKQLNVHWLRSQIG 1147
 T+ALVG+SGCGKSTT+QL++R YDP + +DG D++ +NV +LR IG
 Sbjct: 422 TVALVGNSSGCGKSTTVQLMQRLYDPTGEM-----VSDGQDIRTINVRFLREIIG 471
 Query: 1148 IVSQEPVLFDCSLAENIAYGDNSRSVSMDEIVAAAKAANIHSFIEGLPQVAAVNQKWL I 1207
 +VSQEPVLF ++AENI YG +V+MDEI A K AN + FI LP

Sbjct: 472 VVSQEPVLFATTIAENIRYG--RENVTMDEIEKAVKEANAYDFIMKLPH----- 518
 Query: 1208 PHLIDSHGAAHDHLHHIQTVSEQRDYDTQAGDKGTQLSGGQKQRVAIARAIIRNPKLLLLD 1267
 ++DT G++G QLSGGQKQR+AIARA++RNP+LLLL
 Sbjct: 519 -----KFDTLVGERGAQLSGGQKQRIAIARALVRNPKILLD 555
 Query: 1268 EATSALDTESEKVRLYSHQHCTSCVQVQVEALDQARKGRTICIVVAHRLSTIQNADCIASF 1327
 EATSALDTESE V VQ ALD+ARKGRT IV+AHRLST++NAD IA F
 Sbjct: 556 EATSALDTESEAV-----VQVALDKARKGRTTIVIAHRLSTVRNADVIAGF 601
 1328 QGGVVVEKGTHQQLIAKKGYYHMLVTKQMGYHSGVEADIITSTIKNKS 1375
 GV+VEKG H +L+ +KG+Y LVT Q +G E ++ + ++KS
 Sbjct: 602 DDGVIVEKGNHDELMKEKGIYFKLVTMQT---AGNEVELENAADESKS 646

Score = 152 bits (379), Expect = 7e-36

Identities = 78/124 (62%), Positives = 94/124 (74%), Gaps = 14/124 (11%)

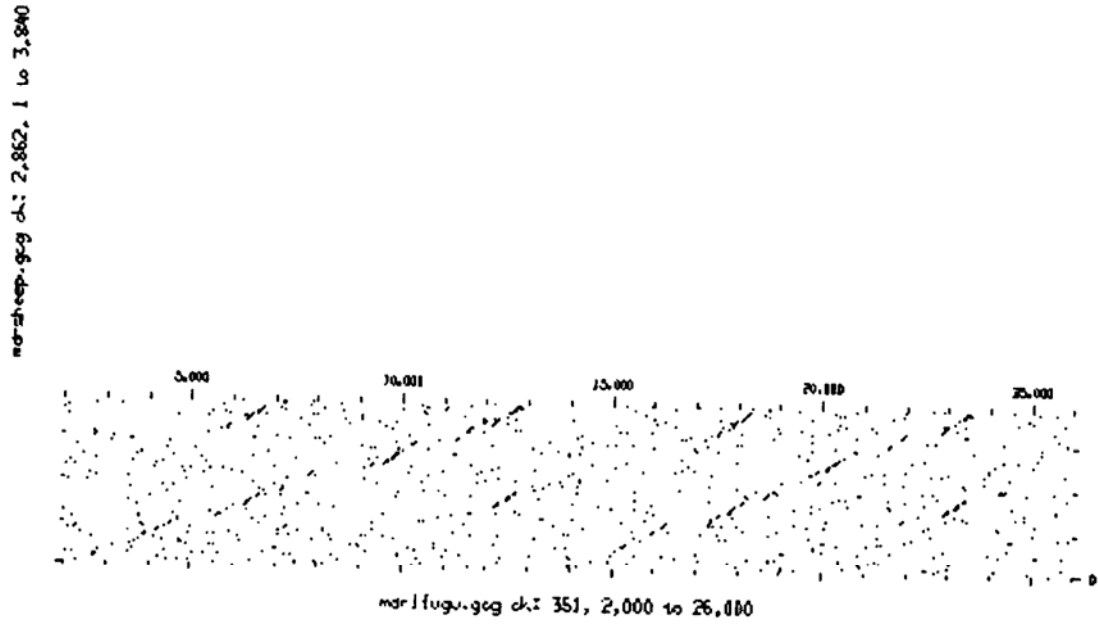
Query: 1 YDTQAGDKGTQLSGGQKQRVAIARAIIRNPKLLLLDEATSALDTESEKVRLYSHQHCTSC 60
 +DT G++G QLSGGQKQR+AIARA++RNP+LLLLDEATSALDTESE V
 Sbjct: 520 FDTLVGERGAQLSGGQKQRIAIARALVRNPKILLDDEATSALDTESEAV----- 568
 Query: 61 VQVQVEALDQARKGRTICIVVAHRLSTIQNADCIASFQGGVVVEKGTHQQLIAKKGYYHML 120
 VQ ALD+ARKGRT IV+AHRLST++NAD IA F GV+VEKG H +L+ +KG+Y L
 Sbjct: 569 ---VQVALDKARKGRTTIVIAHRLSTVRNADVIAGFDDGVIVEKGNHDELMKEKGIYFKL 625
 Query: 121 VTKQ 124
 VT Q
 Sbjct: 626 VTMQ 629

同时我们利用其他生物的多药耐药基因的全长 cDNA 作为纵坐标,发现 contig2 中确实存在与多药耐药基因家族同源的序列。

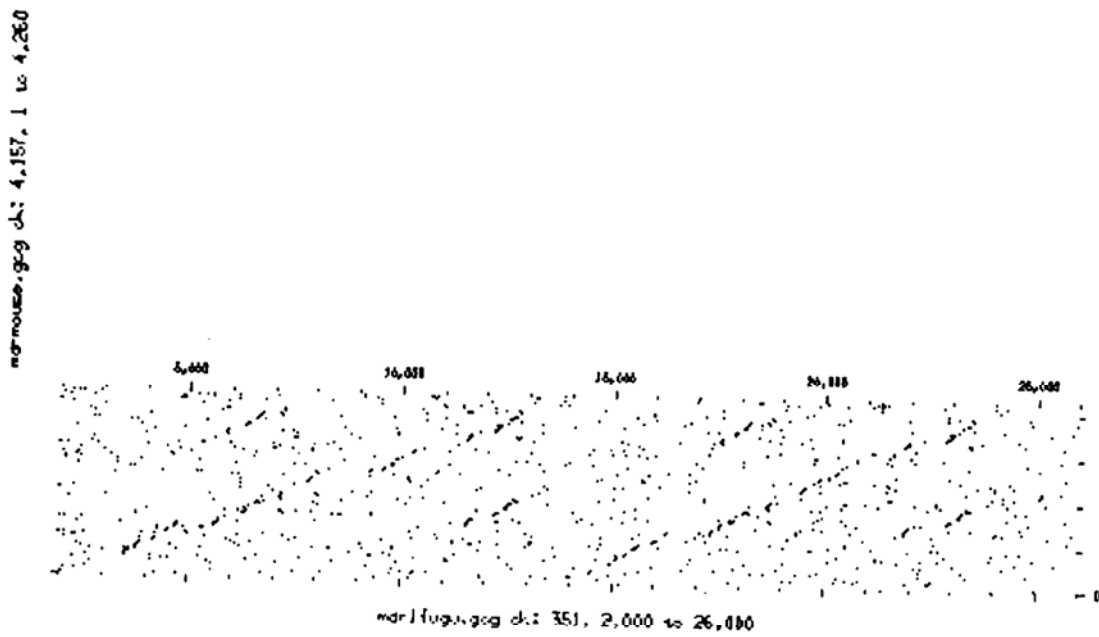
当然对于一未知基因如在数据库没有其同源基因, blastx 将失去意义, 此时利用 Grail 软件对其编码序列进行一初步的预测, 将可达到很好的效果。

blastx 分析的结果只是一段一段的同源序列, 能不能将这一段一段的同源序列“拼凑”成一个完整的 cDNA 呢? 首先根据“ag.....gt”的剪接法则来判断河豚鱼的同源序列部分是否符合该标准。剪接位点预测软件可帮助我们较简单的获得整个基因序列中可能存在的剪接位点(donor 和 acceptor), 一般对于人的基因组序列分析是选定剪接位点的 Score 值>0.4 合适, 考虑到河豚鱼与人之间的巨大差异(例如进化方面)。本文选定 Score>0.001, 结果发现大多数同源序列确实遵守着“ag.....gt”法则, 有少数同源序列错开 1-2bp 也正好符合该法则, 但这种错开的 1-2bp 往往与其前或后也同样必须错开 1-2bp 才能符合该法则的序列相互匹配,

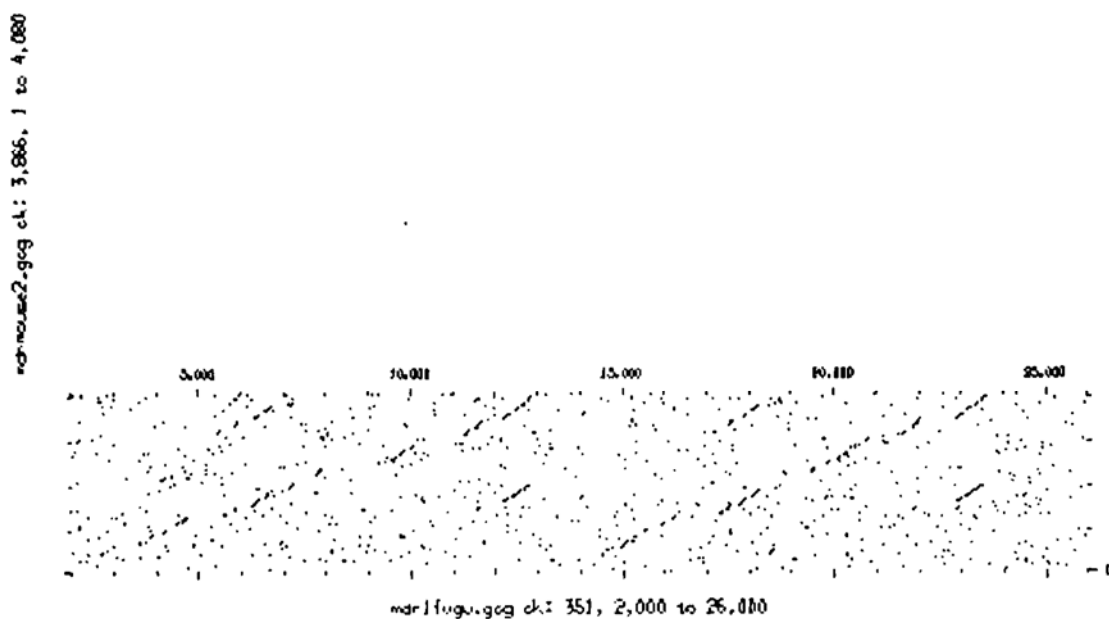
Contig2 与绵羊的多药耐药基因的同源性比较:



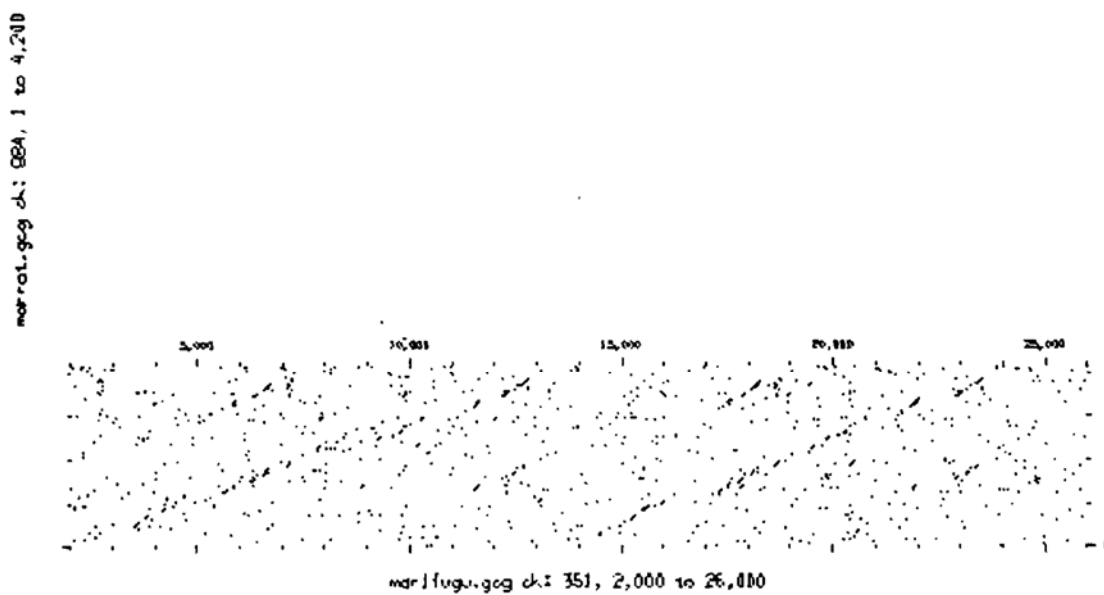
Contig2 与小鼠的I型多药耐药基因的同源性比较:



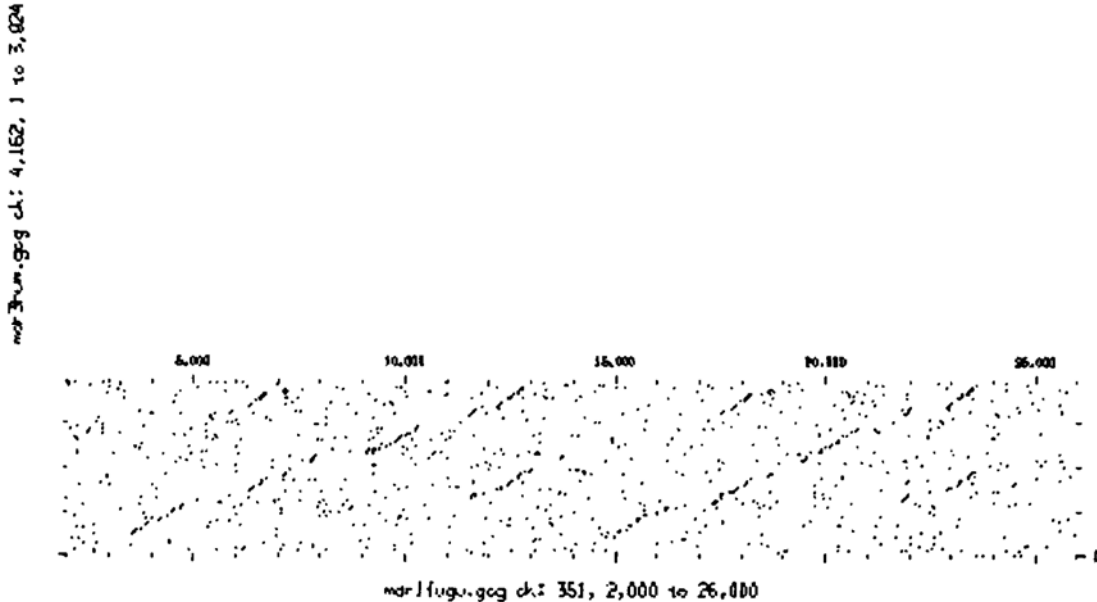
Contig2 与小鼠的 II 型多药耐药基因的同源性比较:



Contig2 与大鼠的多药耐药基因的同源性比较:



Contig2 与人的 III 型多药耐药基因的同源性比较:



也就是说错开 1-2bp 可以通过该序列前后也同样错开而得到补偿, 不至于失去正常的阅读框架, 符合内含子的类型具有三种类型这一现象(详见综述二)。另外, 大多数 “ag.....gt” 位点的 Score 值都很高。

fuguMDR2 的内含子类型与人的 MDR1 的内含子类型完全一致, fuguMDR1 除 intron6 外, 其余的内含子的类型也与人的完全一致, 虽然从这数据中我们并不能获得什么肯定的信息, 但至少说明基因进化中的高度保守性, 似乎有为综述一中有关内含子进化的理论提供了一些有趣的线索或证据。FuguMDR1 和 fuguMDR2 的内含子 10 比人的相应的内含子要大, 似乎违反了河豚鱼内含子较小的原则, 这有可能暗示内含子 10 中还存在着—外显子, 但剪接位点的检测并不支持这一猜测; 如果将来获得的真正的河豚鱼 MDR1 和 MDR2 的内含子 10 确实比人的对应内含子大, 是否该内含子中存在着调控结构呢? 也许这违反常规的现象正是河豚鱼作为模式生物的魅力所在, 换句话说, 比较基因组学只有在一些特殊的差异上才能充分发挥其在基因表达调控研究的作用。

为避免一些同源性很小的序列也同样相互比较, 本文使用 blastx 时采用的相似值=0.0001, 对于人的 MDR1 cDNA 一些外显子未发现同源序列, 本文采用将该区段序列单独列出, 选择相似值=10, 这样我们就获得了 exon4 和 exon16 的同源序列, 但由于 exon16 的同源序列的位置不符合 “ag.....gt” 法则, 故本文采取猜测的方法, 原则有三: 一符合 “ag.....gt” 法则; 二编码序列的长度与人的 exon16

的长度相差不大；三其间不可出现终止密码子。Exon5 的获得也是同样的方法，

FuguMDR1 和 fuguMDR2 的内含子大小及其类型

内含子 序 号	fuguMDR1		fuguMDR2		humanMDR1	
	大小 bp	类型	大小 bp	类型	大小 bp	类型
4	131	1	72	1	>20000	1
5	516	2	231	2	3100	2
6	104	0	103	2	541	2
7	210	0	81	0	4700	0
8	121	2	307	2	7500	2
9	87	0	88	0	2800	0
10	862	0	1013	0	250	0
11	88	0	70	0	170	0
12	326	0	173	0	200	0
13	86	0	86	0	300	0
14	379	0	385	0	3200	0
15	93	0	93	0	850	0
16	216	0	386	0	600	0
17	81	0	106	0	2700	0
18	1121	0	80	0	1900	0
19	275	0	146	0	2700	0
20	75	0	75	0	4800	0
21	103	0	88	0	1100	0
22	99	2	97	2	1200	2
23	875	2	709	2	2500	2
24	89	0	85	0	1100	0
25	453	0	820	0	5400	0
26	104	0	104	0	3300	0
27	74	0	74	0	1400	0

但必须再增加一条原则：增加 1-2bp 而不影响整个阅读框架。根据人的 MDR1 cDNA 的 exon16 和 exon4、exon5 正好是 N 端半侧 C 端半侧的起始序列，其功能非保守性，故而在进化中变异较大，而致与人的同源性较差。其它的一些重要结

构域如跨膜区和核苷酸结合位点结构都表现极强的保守性。对于人的MDR1 cDNA的 exon1 和 exon2、 exon3 未在两段序列中发现同源序列，且这三个外显子编码结构在整个多药耐药基因的功能上并无多大关系，故本文未预测他们在河豚鱼中相对应的位置。

FuguMDR1 和 fuguMDR2 的外显子的碱基数目是一致的（除 exon6、7、16、19）。简单的数字并不能说明什么重要问题，但该有趣的现象的存在似乎有一定的道理，究竟是由于预测的方法的欠缺呢，还是河豚鱼的多药耐药基因本身就是这样。如果真是这样，则河豚鱼的多药耐药基因间在进化上的保守性是显而易见的。另外，这些外显子的碱基数目与人的MDR1 和MDR3 的外显子碱基数目也是完全一致的。如果将来河豚鱼多药耐药基因的表达结果确如预测这样，那该基因在不同物种间进化上的保守性是很清楚的。

promotor 和加尾信号软件的使用，为我们寻找一个完整 ORF 提供了方便，根据 promotor 的位置，我们搜索起始密码子，根据 balstx 结果和加尾信号，搜索终止密码子。

这样，我们就预测到了 2 个完整的河豚鱼多药耐药基因。

对于预测基因结构，目前有很多软件可供我们使用，我们既可通过工作站来完成，又可通过互联网来完成上述工作。本文所采用的软件系统都是作者所能及之物，一些可能更好的软件由于种种原因（如需购买软件或需有偿使用）而不能用来分析基因组序列。

多药耐药基因编码一种膜蛋白，其功能是依靠 ATP 提供能量，将细胞内的有些物质转运到细胞膜外（详见综述二），其结构特点是具有 12 次跨膜结构，N 端半侧 C 端半侧各 6 次，在 N 端半侧的 C 端和 C 端半侧 C 端都存在核苷酸结合位点与 ATP 结合，依靠 ATP 提供能量后发挥转运功能。基因组进化学研究发现：不同生物的多药耐药基因或 ATP 结合蛋白之间都存在着 12 次或 6 次跨膜结构，核苷酸结合位点的氨基酸序列很保守，河豚鱼是否也存在这样的性质呢？作者试图以跨膜结构预测软件来分析预测，但总是获得 11 次跨膜结构，以人的 MDR1 cDNA 为模板，用该软件来分析，也同样只获得 11 次跨膜结构，缺少了第 7 个跨膜结构，这样，作者认为河豚鱼的这两个多药耐药基因也具有 12 次跨膜结构，只是使用的软件系统功能较差而已。通过氨基酸序列比较，我们发现跨膜结构区在不同的生物之间还是比较保守的，但其保守性比不上核苷酸结合位点。

多药耐药基因的 N 端和 C 端具有较强的相似性，由此而引起一系列关于多药耐药基因起源的假说（详见综述二）。FuguMDR1 和 fuguMDR2 是否也具有这样的特征呢？，以 exon15 为分界线，分别将 FuguMDR1 和 fuguMDR2 的 5'端和

3'端分别进行同源性比较,发现河豚鱼同样存在这样的特征。

多药耐药基因作为一个家族,在同一物种中可有多种亚型,人有 2 种亚型,小鼠有 3 种亚型,从本实验结果来看,河豚鱼至少有 3 个多药耐药基因亚型,其中 FuguMDR1 和 fuguMDR2 已如前述,确定无疑,在 1bp-1kb 中确实存在着另一完整的 ORF 的 C 端所应有的一切,这说明在 contig2 的前侧序列(当然不在 cosmid124A22 中),存在着另一 fuguMDR3 的 ORF 的 N 端,据此,作者认为河豚鱼具有三种亚型的多药耐药基因。

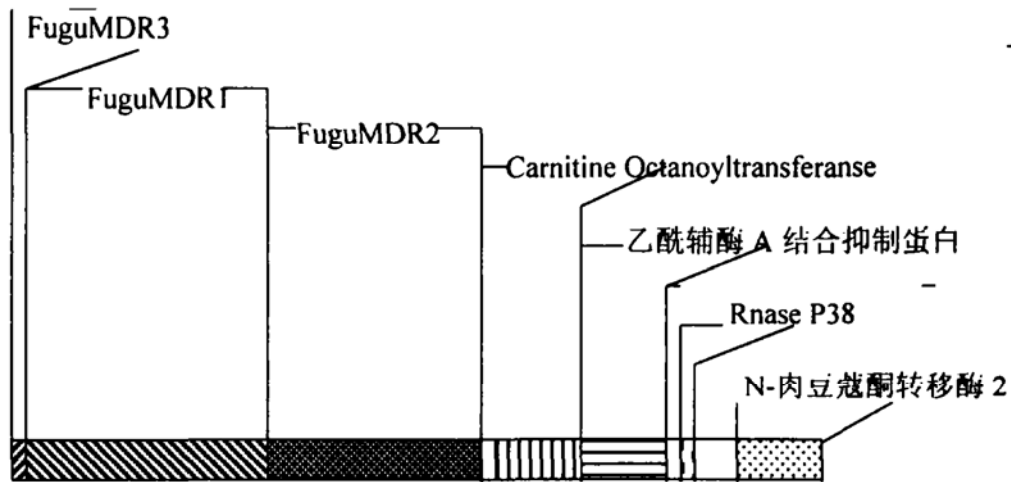
数字统计资料

	fuguMDR1	fuguMDR2	humanMDR1
基因全长	10.787kb	10.223kb	> 120kb
cDNA	3732bp	3699bp	3723bp
G+C%	53.03	55.18	44.43
A+T%	46.97	44.82	55.57
氨基酸(exon4-28)	1244aa	1233aa	1241aa
外显子数日	24	24	24
内含子数日	24	24	24
分子量	138.2kd	136.1kd	137.1kd
Strong Basic(+)			
Amino acids(K,R)	122	124	137
Strong Acidic(-)			
Amino acids(D,E)	130	127	125
Hydrophobic aa			
(A,I,L,F,W,V)	499	495	511
Polar aa			
(N,C,Q,S,T,Y)	322	319	296
等电点	6.549	6.977	8.822
charge at pH7.0	-4.865	-0.180	14.966

这些数字都表明河豚鱼 fuguMDR1 和 fuguMDR2 虽然和人的 MDR1 cDNA 序列相似,但其基因全长要远远小于人的基因组, fuguMDR1 只有人的 1/11, fuguMDR2 只有人的 1/12, 比一般总的规律:河豚鱼的基因组只有人的基因组的 1/7.5 还要小,这样本课题就达到了所要设想的目的:河豚鱼存在着与人同源的基

因, 这种同源不仅表现在序列上, 包括核苷酸和氨基酸序列的同源, 也同样表现在功能的同源上, 这为河豚鱼作为模式生物提供了有力的证据。

在 cosmid 124A22 剩余的 14kb 基因序列中, 通过 blastx 搜索, 找到 4 个与人有着同源性的基因, 这些基因并不同于河豚鱼多药耐药基因的三个类型, 后者是以头尾相连方式相连, 而前者却有尾头相连的方式。通过下述简式图, 我们可容易看出 cosmid 124A22 中六个基因的排列和转录顺序。



Cosmid 124A22 中 6.5 个基因的分布模式图

在 cosmid 124A22 中基因的密度为 $40\text{kb}/6.5=6\text{kb}$, 基本符合文献报道 7-8kb 的基因密度。这表明河豚鱼作为模式生物不愧为人类基因组计划的一个理想“沙盘”模型。

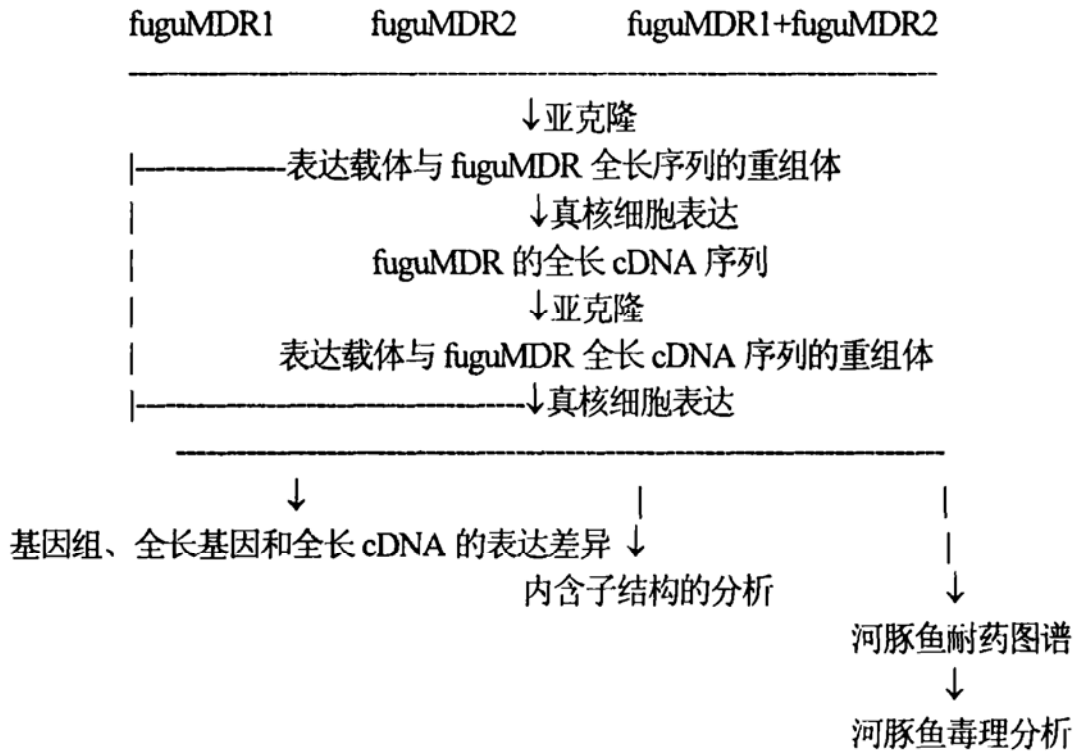
人的 carnitine acetyltransferase 位于人的 9q34.1^[6], 乙酰辅酶 A 结合抑制蛋白位于人的 2q12-21^[7], 其他两种基因未能找到其在人的染色体上的相应位置。表面上看来, 这六个基因似乎在进化上并不符合同线保守或“基因板块”学说, 但一般情况下, 同线保守或作为板块遗传的基因一般认为相互之间存在相互调控的因素, 这种调控表现为各保守区内的基因功能之间的相互依赖性。从这些基因的功能来看, 彼此之间并无多大关系。但注意到河豚鱼的多药耐药基因的三个亚型之间是绝对保守的, 彼此之间头尾相连, 与人的 MDR1 和 MDR3 是完全一致的。这样我们可以猜测: 多药耐药基因是作为一个板块或同线保守区而得以在各生物之间保守遗传。三种多药耐药基因之间可能存在着相互调控的因素而并不仅仅限于基因内部的调控。并不是所有的多药耐药基因都能正常表达这一现象, 如人的 MDR3 的存在但不表达似乎说明多药耐药基因的各亚型与生物体多药耐药的表型有着千丝万缕的联系。河豚鱼小而紧凑的基因组结构正好为研究基因之间的调控

提供极佳的条件,这将在下文阐述,而另4种基因是作为相互独立遗传的基因存在,可能彼此之间并不存在相互调控的作用,故而不表现出同线保守性。

河豚鱼作为模式生物的意义不仅仅表现或应用在比较基因组学,它也广泛应用于表达、调控的研究方面,也应用在致病基因的克隆方面。河豚鱼作为模式生物应用在基因克隆方面的具体策略在综述一中已有详细阐述,本实验室正以此策略克隆5p猫叫综合征的致病相关基因。

河豚鱼作为模式生物在研究基因的表达、调控方面的研究正在起步,仅仅通过比较基因组学尚不足以完全解决基因的调控问题。由于人的基因的全长序列很难获得,更谈不上基因组的分离。本文的序列结果已经显示出,我们可以很容易的获得fuguMDR1和fuguMDR2的全长序列,也可以获得二者的全长基因之和,如果我们将这三段序列分别克隆至表达载体中,通过直接表达后获得二者的全长cDNA序列,有了真正的全长cDNA后,可以找到真正的阅读框架,从而确定各内含子与人或其他动物对应内含子之间的同源保守,以根据“功能越重要,同源性越高”的原理来寻找存在于内含子中的重要功能区,然后通过基因组、全长基因、全长cDNA的表达差异来寻找基因的调控结构。另外,可通过研究河豚鱼多药耐药基因的耐药谱,进而与其他生物的耐药谱相比较,也许可以发现多药耐药基因表达产物与底物结合的一些有趣现象。

本课题的延伸简示图



综上所述, 通过异种杂交我们可以获得河豚鱼与人同源的基因, 河豚鱼的基因组很小, 只有人的 1/7.5, 基因密度很大。在 cosmid 124A22 的 40kb 序列中含有 6.5 个基因, 其中包括 2.5 个多药耐药基因, 河豚鱼作为一新型模式生物必将在人类基因组中发挥越来越重要的作用。河豚鱼多药耐药基因基因组的分离、克隆、序列分析, 必将会为研究多药耐药基因的功能、表达调控提供新的方法、线索。

参 考 文 献

1. Hoheisel.J. (1992) *Anal.Biochem.*209.238
2. Erase-a-Base. System Technical Manual,#TM006, Promega Corporation.
3. Steven Henikoff. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA Sequencing. *Gene*,28(1984)351-359.
4. Henikoff.S. Ordered deletions for DNA sequencing and in vitro mutagenesis by polymerase extension and exonuclease III gapping of circular templates. *Nuceic.Acids.Res* 1990,May 25;18(10)2961-2966.
5. Ahemd A. Isolation of nested,bidirectional deletions by the use of exonuclease III on double-origin vector. *Gene* 1994 Dec.30;151(1-2):97-101
6. Corti O,Finocchiaro G, Rossi E, Zuffardi O, DiDonato S, Molecular cloning of cDNAs encoding Human carnitine acetyltransferase and mapping of the corresponding gene to chromosome 9q34.1.*Genomics* 1994 Sep 1;23(1):94-99
7. DeBernardi MA,Crowe RR, Mocchetti I, etal, Chromosome localization of the huamn diazepam binding inhibitor gene. *Pro.Natl.Acad. Sci USA* 1988 Sep;85(17):6561-6565.

综述一

河豚鱼作为新型模式生物在人类基因组研究中的应用

人类在认识和改造自然的过程中,总或多或少的需要借助于其他一些工具。正是这些工具的出现,才进一步完善了人类的认识论;同时新的认识又产生了新的工具。如此良性循环,发展至今天的知识爆炸年代。人类在认识自我(这里主要指躯体)的过程中,也同样是这样的道理。在认识生、老、病、死的道路上,经历了二次质的飞跃,即三次主要阶段:第一阶段是利用天然物质来治病和保健;第二阶段是利用仿生药物及近代产生的病原学、病因学来诊断、治疗疾病,这两个阶段都属于被动认识阶段,且只限于对疾病的认识,对于人的生、老、死的本质则一无所知;第三阶段是随着遗传学、分子生物学等学科的建立与发展,人类才终于找到了决定自己一生的“天书”——人类基因组,真正知道了无论是生还是死、无论是老还是病都是由人的基因决定的,认识和改变基因将是最终改变人的生与死、病与老的唯一途径。由此而产生了人类基因组计划。该计划的近期目标是完成四张图——遗传图、物理图、序列图和转录图的绘制^[1],它的最终目的,也是一个无截止日期的目标是找出人类基因组所有的基因以及这些基因的结构、功能和表达、调控。为了达到这些目的,一个特殊的辅助工具由此而产生——模式生物。好的工具的选择,将达到事半功倍的效果。

新型模式生物的选择

现已被人类基因组计划正式接受的模式生物有:酵母^[2]、线虫^[3]、果蝇^[4]和小鼠。它们所起的作用其他作为非模式生物的工具所无法替代的。

酵母作为最简单的真核生物,一直是作为研究真核生物基因(主要是单基因)的结构、功能和表达、调控的有效模型;

线虫由于具有躯体透明、结构简单(为简单的多细胞生物)的特点而一直成为研究细胞分化和组织发育的最理想模型;

果蝇作为经典遗传学的“功臣”,不仅仅与遗传连锁率的确立有关,由于其特殊的染色体结构,又使其与细胞遗传学中多线染色体的发现以及分子遗传学中

同源盒的发现密不可分。

小鼠与人同属哺乳类，由于其繁殖周期短、经济易得、许多生理和病理现象与人类类似，使其一直成为研究遗传、病理、生理、药理的最佳模型。

其中酵母、线虫、果蝇的基因组大小分别为 14Mb (megabase)^[5]、100Mb 和 165Mb，远较人类基因组 3000Mb 小得多，对它们的基因组作图，在很大程度上起到人类基因组作图的参照系的作用，但它们又有其不足之处：酵母、线虫、果蝇与人类的进化距离相距甚远^[6]（图 1），其基因组无论是结构还是功能、更多的是基因板块的保守性与人类基因组差异太大；小鼠与人的进化距离很近，但其基因组与人类的一样复杂。由于人类基因组 94% 的区域都为基因间序列和内含子序列，这些序列尤其是其中的大量的重复序列的存在给实际工作造成了诸多困难：

1. 使得基因组的各亚克隆之间很难组成“克隆重叠群”，同时，在完成序列图的过程中，重复序列的存在造成了大量的无效劳动和财力浪费；2. 单拷贝序列很难在整个基因组中准确定向定位，分离表达序列工作（如致病基因克隆）显得十分艰难，犹如大海捞针；3. 在研究基因的表达、调控的过程中，一方面会由于庞大的内含子序列而使转基因工作（包括细胞转染和转基因动物）难度加大甚至不可为，另一方面会因缺少存在于内含子中的一些增强子及其一些尚未知晓的内含子特殊功能的作用，而使单纯利用 cDNA 来进行研究的过程“失真”，更谈不上进行基因组功能的研究了。

这样，在完善现有的一些技术、设备之外，寻找一种新型模式生物成为人类基因组计划中迫在眉睫的任务之一：这种模式生物是人类基因组的“作战沙盘”，该“沙盘”具有以下特征，①各独立单位基本存在，且去粗存精，性质基本不变；②各单位之间的相对关系（包括位置、作用）基本不变。组成基因组的各单位就是散布于人的 24 条染色体的近 100000 个基因，基因内的独立单位就是一个个外显子，其性质主要指基因的功能，各单位之间的相对关系就是指基因之间的调控及空间关系。

回顾生物界脊索动物的进化过程（图 2）^[7]，最高等的哺乳动物—人类是从硬骨鱼类的内鼻类（Sarcopterygii，又称肉鳍类）中的总鳍鱼的一支分化出来的。作为最原始的脊椎动物—硬骨鱼，是否与最高等的脊椎动物—人类之间在基因组方面有许多共性呢？也就是说，能否从鱼类中找出人类的“基因沙盘”呢？

鱼类基因组的特性^[8]：

1. 基因组大小：一般鱼类的基因组大小位于 0.5—4.5pg 之间^[9]，大部分分布于 1pg 左右（图 3）。鱼类中一般基因组的大小与其染色体的数目呈正相关关系，这在其他生物中，几乎不见此种现象。

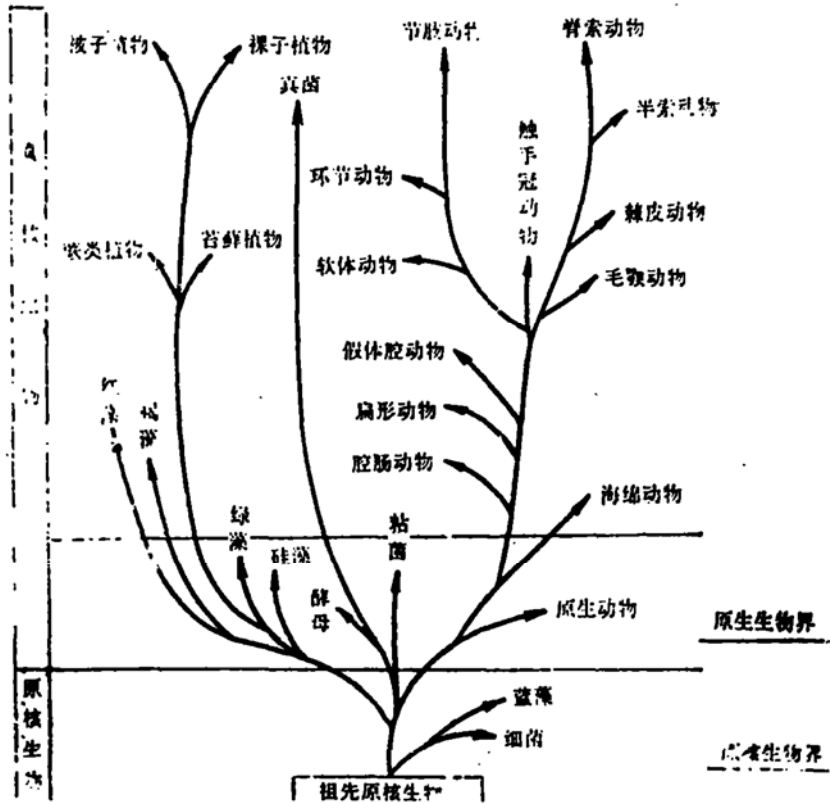


图 1. 生物进化图及相关模式生物的位置

2. 染色体的数目：通过对 1000 种鱼类的核型分析，鱼类的染色体数一般为 8-86 对^[10]，80%的为 20-28 对，这其中约一半为 24 对。这表明所有脊椎动物的原始染色体数目可能都为 24 对^[11]。

3. 基因的构成：通过密度梯度离心的方法发现，鱼类的基因组都为 G、C 稀疏区 (L1、L2) (图 4)^[12]，GC 水平含量与 DNA 片段中基因所占的比例呈正相关关系。在人类中，有三条 GC 富含区 (H1、H2、H3)，其中 H3 只占基因组的 5%，但大部分人的基因都位于 H3 中，H3 一般位于染色体端粒^[13]。GC 富含区对应于细胞遗传学中的染色体 R 带，GC 稀疏区对应于 G 带，我们将此关系称之为等容线 (isochores)，等容线的出现架起了基因与染色体的桥梁。由此推测：鱼类的基因在基因组中呈现均匀分布状态，而不象人类的基因在染色体上呈极性分布。

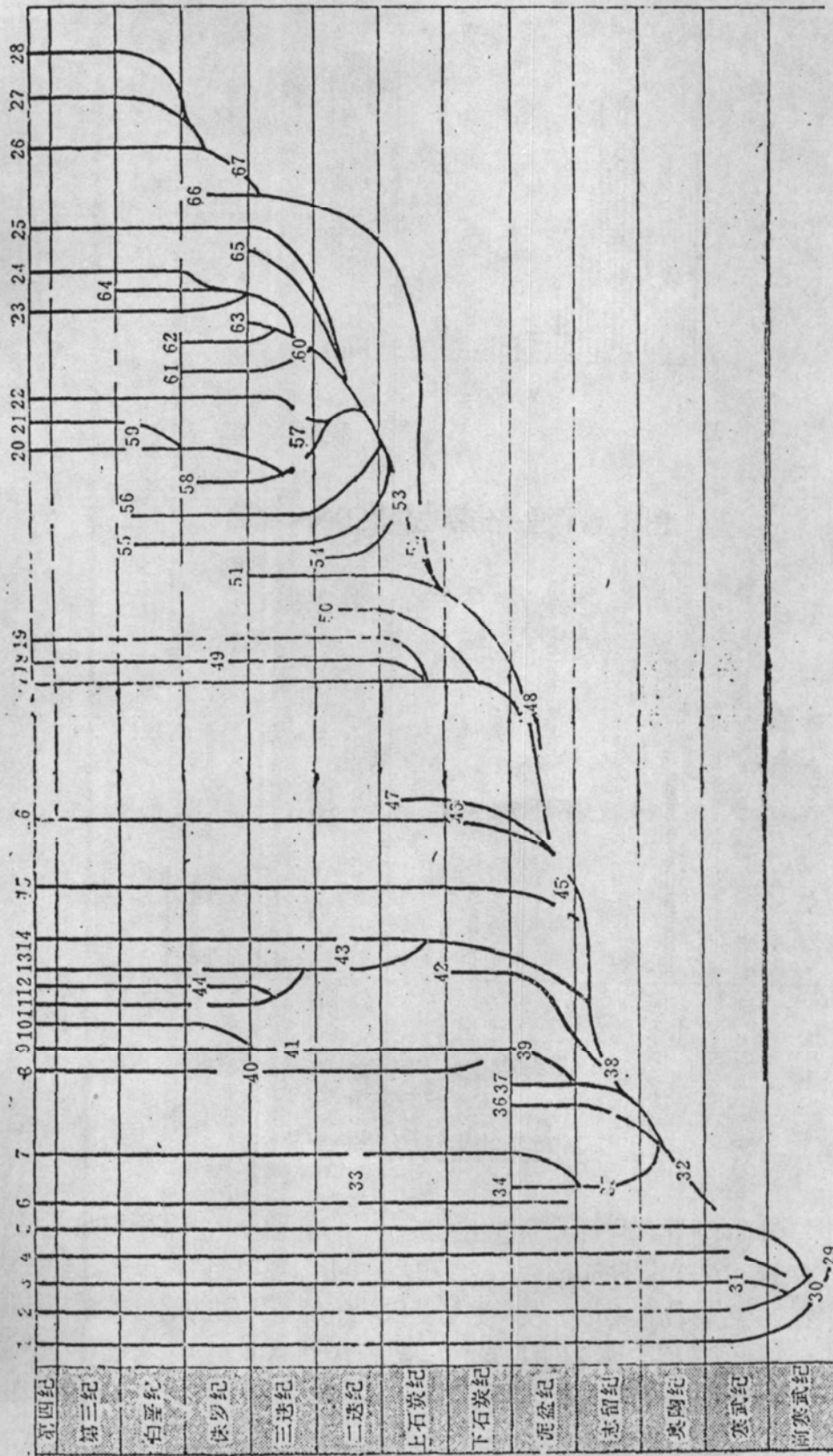


图 2 脊索动物在地质时期的谱系关系系 (自 Klinge, 1977)

- 1. 非脊索动物; 2. 海鞘纲; 3. 樽形鞘纲; 4. 尾樽鞘纲; 5. 头索动物; 6. 盲鳗; 7. 七鳃鳗; 8. 根鳗; 9. 鳗目; 10. 鳗目; 11. 鳗目; 12. 鳗类; 13. 早期的金骨类; 14. 软骨硬鳞类; 15. 肺鱼; 16. 鳗类; 17. 无尾类; 18. 有尾类; 19. 无尾类; 20. 鳗类; 21. 蛇类; 22. 楔齿蜥; 23. 龟鳖类; 24. 鸟类; 25. 龟鳖类; 26. 原兽类; 27. 后兽类; 28. 真兽类; 29. 棘皮动物祖先; 30. 脊索动物; 31. 尾索动物; 32. 脊索动物; 33. 圆口类; 34. 单系孔类; 35. 无颌类; 41. 板鳃类; 42. 棘鱼类; 43. 棘鱼类; 44. 新鳍类; 45. 内鼻孔类; 51. 迷齿类; 52. 爬行类; 53. 无孔类; 54. 中龙类; 55. 无孔类; 56. 鱼龙类; 57. 鲸龙类; 58. 始鲸类; 59. 有鳞类; 60. 初龙类; 61. 鸟龙类; 62. 翼龙类; 63. 槽齿类; 64. 鲸龙类; 65. 杯龙类; 66. 下孔类; 67. 哺乳类。

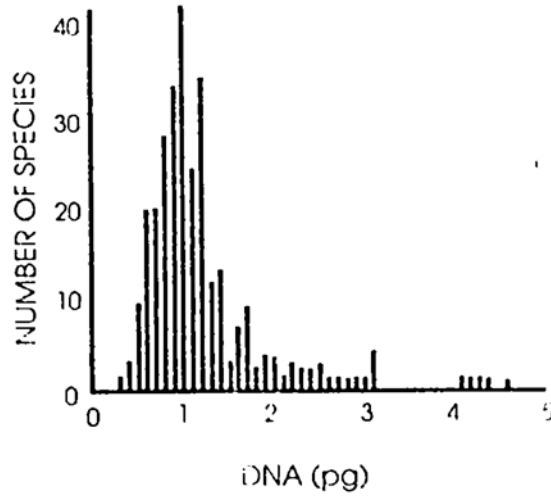


图 3. 硬骨鱼单倍染色体的 DNA 含量

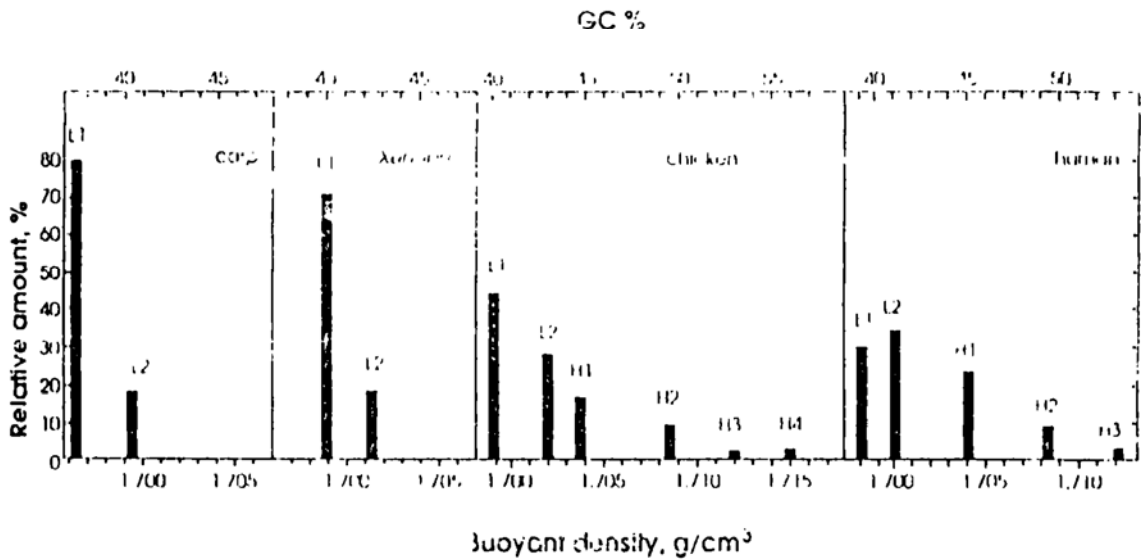


图 4. 不同种类的生物的 GC 含量分布

4. 基因连锁群的进化保守性：在进化的过程中，相连的一些基因彼此之间组成“板块”，板块相对稳定的向下传递，再相互之间组成染色体。虽然这些板块在不同的物种的染色体位置不尽一致，但这些板块中的基因间的线形关系与结构却高度相似，这种现象称之为“同线保守”（conserved synteny）。通过对一些连锁群在不同物种间的分析（表一）^[14]，我们可初步看出“同线保守”的客观存在。

表一：不同种群间的同线保守分析（起源于共同的染色体）
 (Derived from duplication of a single chromosome of a vertebrate ancestor)
 (PEP-D 和 GPI 以及 PEP-B 和 LDH-B)

Poeciliid fish	Salmonid fish	Frogs (<i>Rana</i>)	Mouse	Mouse isozyme	Human
II	13		1		2
GPI2	MDH	HK2	IDH1	LDHA	ACP1
IDH-M	IDH	MANA			IDH-S
LDHA	GPI	PEPB	2	HEXA	MDH1
IDHC	LDHA	LDHB	ACF2	PKM2	10
PK2	LDHC	MPI	SORD	MPI	GOT-S
ENO2	COT		9	NP	HK
MPI	MPI		PK3	CKBB	11
III	3	GPI	MPI	SORD	ACP2
GUK2	GPI	GOT2(M?)	7		HBB
GAPD1	PEPD	PEPD	PEP-D	PEPB	LDHA
ME	SORD	GOT1(S?)	GPI	LDHB	12
IV	7	TPI	IDH-M	CS	PEP-B
PK1	LDHB	IDH1	LDHA	TPI	LDHL
GPI1	PEPB	HB	8	GAPD	GS
IDH1	1		GOT-M	ENO2	ENO2
GOT-M	GOT		19		GAPD
PEPD	MDH		GOT-S	GPI	TPI
V			6	PEPD	1
MDH1			LDHB		NP
GLYDP1			GAPD		CKBB
VI			TPI		15
NP2			14		IDH-M
GLNS			NP		MANA
TF			11		PKM2
UMPCK			HEBA		HEXA
GUK3			10		SORD
AMY			CS		MPI
VII			HK		16
GALT2			APK		GOT-M
IDH2					HBA
U3					19
GOT-S					GPI
MDH1					PEPD
U4					MANB
HEPA					CKBB
TPI					
GAT2B					

Morizot 认为同线保守的原因有：1. 由于功能和/或位置的进化约束，在强烈的自然选择压力下而使功能特异的基因连锁群得以保存；2. 脊椎动物常用的基因复制机制，导致了一个或更多的特异性的有丝分裂机制，这些机制要求染色体保持特异性的基因顺序 (gene arrangements)。

鱼类作为模式生物的优点（遗传和发育学方面）^④：

1. 既能野外繁殖，又能实验室繁殖；
2. 保持着很强的父系信息(husbandry information)；
3. 一个生殖周期能产生几百个甚至几千个卵，这些卵都为受精状态，且很容

易发育成成鱼；

4. 各类鱼往往有许多高度同源的种群，且较容易获得新的突变种；
5. 同一基因编码的蛋白质在不同种的鱼类之间所起的形态学和生理学作用不尽相同，显示了基因功能、表达的多态性；
6. 鱼类作为非恒温动物，能在各种各样的环境中生存，诸如温度、压力、氧饱和度、PH、盐浓度等这些环境因子都会给遗传因子所决定的代谢、生理和行为活动带来极大的改变^[5]。

由此，初步认定可以从鱼类中寻找我们所需要的新型模式动物。

新型模式生物—河豚鱼的建立

1968年，Ralph.Hinegardner^[6]分析了100多种硬骨鱼的细胞DNA总量，发现豚科动物，包括河豚鱼含有较小的基因组；80年代，Pizon^[17]等又对其进行了验证，结果显示河豚鱼的基因组含量为0.4-0.5pg（单倍体），约相当于400-500Mb。

河豚鱼属于脊索动物门，硬骨鱼纲，辐鳍亚纲，全骨总目，豚形目，豚亚目，豚科，东方豚属，与人的进化距离约450万年。“河豚鱼体裸露无鳞或有细弱小刺，颌齿愈合成一片，有明显的中缝；无第一背鳍，腹鳍也无，有气囊，充气时身体膨胀；种类多，分类较困难，常以体色、侧线的有无、鼻孔数目来区分。我国及日本沿海和淡水河流中均有；其肉味鲜美，但内脏、生殖腺及血液含有毒素（即河豚毒素）”。常见的有虫纹东方豚（*Fugu Vermicularis*）等^[8]。

1995年，Kadoo Miyaki^[9]等人对6种河豚鱼的核型进行了分析，确认它们彼此之间都有相同的染色体数（ $2n=44$ ）（图5），新臂数（new arm numbers）均为46。

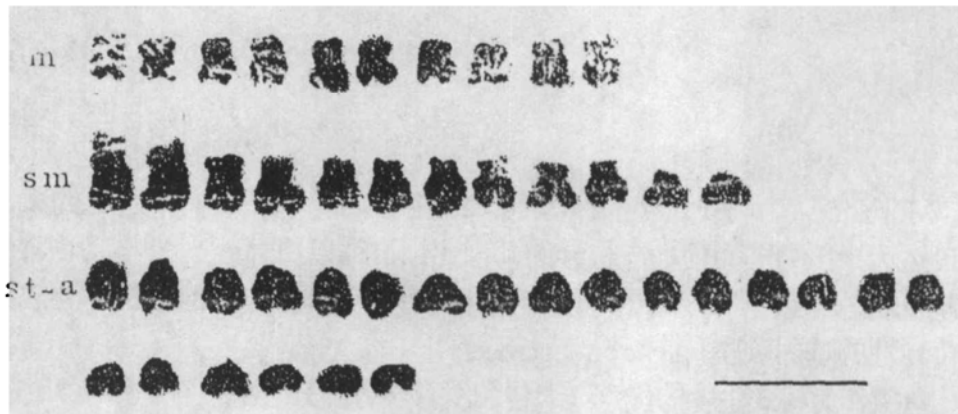


图5. 河豚鱼基因组染色体图（人的染色体图从略）

一、小而全的基因组:

1993年,剑桥大学的Brenner^[20]系统的研究了河豚鱼基因组的特点:

1.通过Southern杂交发现在平均24625个克隆中可含有任何一个基因,按每个克隆平均含有16.5Kb的插入片段计算,河豚鱼基因组大小为 $24625 \times 16.5\text{Kb}=404\text{Mb}$;

2.通过河豚鱼596个克隆测序结果分析,发现所测的127831个碱基中有1011个碱基为编码序列(通过与Genebank比较得知,这又称为电子杂交)。由此推断河豚鱼基因组中编码序列的比例是 $1011/127831=0.791\%$,而人的比例为 0.103% ,这说明河豚鱼基因组中基因密度为人的 $0.791/0.103=7.68$ 倍。如果河豚鱼与人具有相同的基因数,那河豚鱼的基因组大小即为 $3000\text{Mb}/7.68=390\text{Mb}$,这都与Chuou-Houki^[21]等验证的硬骨鱼类的基因组为380Mb相一致,一般现将河豚鱼的基因组大小定为400Mb。在127831个碱基中,重复序列为9762bp,占7.6%,且都成串排列,缺少Alu或LINE那样在人类基因组中散在分布的重复序列。这就是说其中92.4%的基因组为单一序列,而人类只有40%为单一序列;

3.人类基因组中有3%的序列被转录,按每个转录本平均为1000来算,人类共有 $3\% \times 3 \times 10^9 \text{bp} / 1000 \text{bp} = 90000$ 个基因;根据河豚鱼文库某一cosmid全长序列(约40Kb)中含有5773bp的可编码碱基来计算,河豚鱼基因组约含 $5773 \text{bp} / 40 \text{Kb} \times (4 \times 10^8) / 1000 \text{bp} = 60000$ 个基因,虽然计算方法不尽准确,只是粗略估计,但总的趋势说明河豚鱼与人有着基本相同的基因数目^[22];

4.河豚鱼基因组的G+C%为44.2%,没有明显的G带(G、C稀疏区)和R带(G、C富含区),这也说明河豚鱼的所有基因在基因组中均匀分布。

二、单个基因编码结构具有较高同源性与基因组中存在同线保守:

与人对应的河豚鱼部分基因的序列分析结果(表二)^[32]及其他数据显示:1.河豚鱼具有与人一致的外显子及内含子数目且同源性较高,一般认为越是编码蛋白质功能关键区域的外显子其保守性越大,相反其同源性较低;2.河豚鱼内含子普遍较小,一般为60-150bp,没有<50bp的,这可能暗示着50bp可能是脊椎动物内含子碱基数的最低阈值。一般每个基因中都含有一较大的内含子,对应于人同源基因中最大的内含子;3.内含子隐含剪接位点高度保守,即遵守gt.....ag法则^[33](这一结果同样为Brenner的电子杂交结果所证实);4.和其他脊椎动物一样,河豚鱼基因的密码子的第三位碱基一般为G或C,约70%的密码子是以甘氨酸和色氨酸结束^[34],这就为设计兼并寡核苷酸探针来分离多基因家族(multiple-gene family)提供了理论基础和先决条件。

对基因组的研究中发现:

1. 基因间的序列短,基因与基因之间呈高度浓缩状,如Trower^[35]等发现一个40Kb的cosmid中含有8个基因,其中3个基因的总长为12.4Kb,而

与其对应的人 14q24.3 的同源基因组却跨约 600Kb;

表二: 人与河豚鱼部分同源基因的大小比较

基因	大小(Kb)		内含子	
	人	河豚鱼	人	河豚鱼
TSC2	50	17	41	42 ^[23]
PKD1	50(36)	36	46	52 ^[24]
PAH	90(37)	8	12	12
P55	20-30	5.5	?	12 ^[25]
DLST	17	2.2	12	12 ^l
HD	170	23	67	67 ^[26]
V-tRNA-S	21	8.5	?	29 ^[27]
β -cyto-A	3.4	4	5	5 ^[28]
α -skel-A	2.8	3	6	6 ^[28]
Utrophin	~900(39)	40-60	?	>50 ^l
C9	~90(40)	3.9	10	10 ^[29]
G6PD	14	3.8	13	? ^[30]
PP2A α	25.8	3.94	6	6 ^[31]
PP2A β	28.7	4.01	6	7 ^[31]
PPx	8.5	2.21	4	7 ^[31]

2. 基因之间呈一定的保守性, 即同线保守。同线保守广泛存在于与人进化距离较近的小鼠^[36]、牛^[37]、猪^[38]、鸡^[39]等动物中。对河豚鱼部分较小区域的基因连锁分析, 河豚鱼存在着较高程度的同线保守(表三)。一般推测, 由于河豚鱼与人之间的进化距离较远, 连锁打破率(linkage disruption)较高(按照哺乳动物连锁打破率 1 次/2-3 万年计算^[40], 450 万年的进化时间里, 河豚鱼约有 180 次连锁打破, 其平均连锁距离为 400Mb/(180+22)=2Mb), 所以大片段区域的同线保守效果将会下降, 甚至较小区域的基因顺序与方向也不尽完全一致, 但这并不影响其总的同线保守趋势。

另外, 河豚鱼基因组的甲基化方式和 CpG 二核苷酸的分布也与哺乳动物有着显著的区别。其 CpG 二核苷酸高度甲基化, 期望值为 0.6, 高于哺乳类的 0.2, 但低于果蝇的 1.0, CpG 岛少见于基因 5'端而随机分布^[34]。

表三: 河豚鱼部分小区域同线保守序列

基因位置 (河豚鱼)	大小范围	基因位置 (人)
G6PD	all in one 60Kb contig	Xq28
Glycerol-6PD		12
Adenylate cyclase VI		12q12-q13
L-type calcium channel polypeptide		12q13
P55	Both in one cosmid	Xq28
GABA alpha receptor		Xq28
Cannabis receptor	Both in one cosmid	6q14-q15
GABA rho receptor		6q14-q21
Tyrosine hydroxylase	within 10Kb of each other	11p11.5
Insulin-like growth factor receptor II		11p11.5
Phenylalanine hydroxylase	within 20KB of each other	12q22
Insulin-like growth factor receptor I		12q22-24
TSC2	in one cosmid	16q13.3 ^[41]
ADPKD1		16q13.3
SSTR5		16q13.3
Fos	all in one cosmid	14q24.3 ¹
s31		14q24.3
s20i15		14q24.3
7SL		
Activating factor 3		
Fos-like gene		
Dihydrolipoamide succinyl transferase		
Hoxb cluster(9 genes)	90Kb	17q21-22 ^[42]
Hoxc cluster(9 genes)	66Kb	12q12-q13
IgVH genes(>6 genes)	all in one cosmid	14q32.3
MAP-2	all on one cosmid	2q32-q35 ^[43]
MYL-1		
CPSIII		
Isotocin(IT)	all on one cosmid	tail-to-tail ^[44]
Vasotocin		tail-to-tail
Other five genes		
C9	all on one cosmid	5p13

DOS2		5p13
GAS-1		9q22
FBP		9q22
PDGFR	on a cosmid	2.2Kb distance ^[45]
CSF1R		

河豚鱼作为模式生物的研究和应用价值

由于河豚鱼基因组具有以上特点, 其可作为一新型模式生物发挥以下积极作用:

一、人类基因组计划的参照系:

根据其基因组相对于人来说, 既“全”又“小”, 可与其他模式生物的基因组序列一样作为人类基因组计划中四张图的又一参照系, 这既可避免因大量重复序列的存在而很难搭建“克隆重叠群”的尴尬, 又可省钱、省时、省力。

二、为研究基因的结构、功能和表达、调控找到一简便的坐标:

1. 内含子的研究: 生物界一直对内含子具有怎样的功能的问题至今未有定论。内含子的存在肯定有其积极的意义, 即其具有一定的功能是公认的, 如有些内含子起着增强子的作用或一个基因的内含子可能为另一基因的外显子等等。从内含子中发现基因表达的增强子已成为比较基因学最具有特色的贡献之一。因为增强子是否存在、存在于哪一个内含子中, 我们无法通过构建不同的表达质粒(即 cDNA 与任一内含子的不同组合)来验证。通过对不同物种的同源基因的比较, 我们可直观地看到内含子在进化过程中的趋势, 根据“越保守可能越有作用”这一原则来初步找到可能会起作用的“候选内含子”, 进行更深层次的研究, 从而起到事半功倍的效果。

2. 基因重要结构域的发现: 一个基因的所有外显子编码的蛋白质中的所有结构域并不是都起着同样的作用。以前大多数依靠对蛋白质的结构、功能分析, 如抗原抗体反应或通过突变体的功能比较等途径来发现哪些结构域对蛋白质的功能起着决定性的作用。这些方法既费钱、费力, 又费时。根据上述“越保守越重要”的原则, 通过对河豚鱼和人类之间的同源基因结构的电子杂交, 同源性越高的外显子所起的作用可能越大, 同源性越小的作用越小^[46], 如 Maheshwar 对 TSC2^[47] 基因在人和河豚鱼中进行了比较研究, 发现了功能重要的结构域确实有高度的保守性, 还由此发现了该基因另外三个高度保守的区域, 可能也具有重要的功能, 这当然有待于进一步的研究。但是由于小鼠虽具有普遍高的同源性但基因组与人一样复杂, 而线虫、果蝇虽具有很小的基因组但又都具有普遍低的同源性, 使它们失去了这一重要作用。

3. 基因的表达、调控的研究更趋“自然化”：人的基因普遍较大，由于转基因技术条件的限制，人们通常采用的是利用 cDNA 来研究基因的表达，根本谈不上对基因组实施基因之间相互调控的研究了。这种趋向“简单化”的研究方式必须给实验结果的“原始、真实、可靠”性大打折扣。河豚鱼基因的“压缩”与基因之间的“去粗存精”使得在现有技术条件下对单个完整基因甚至是一基因组的转基因工作成为现实，从而实现了基因表达、调控的“自然状态化”，对实验结果产生的影响不言而喻。

上述作用必须基于同源基因的功能不变，Venkatesh B 等人通过转基因小鼠成功的实现了这一设想^[48]。

三、环境因子对基因结构、功能及表达、调控的影响研究的理想模型：

基因病的概念已逐渐取代遗传病而为众多的科学家和世人所接受，这就是说，各种疾病都直接或间接与基因有关。对于各种单基因病、多基因病及部分获得性基因病，人们已想出了许许多多的办法来加以系统研究，而对于环境因子对机体健康的影响，就是环境因子对基因的结构、功能和表达、调控的影响的研究，目前还未见良策。如果说先天性遗传病和病原微生物感染而致的获得性基因病是影响人类健康的大敌，那环境因子对人类健康的影响将与之不相伯仲。随着社会文明程度的提高和工业化水平的增长，各种环境因子的改变对人类健康的影响已越来越引起各国政府及科学家的广泛重视。各种伦理、道德、法律、社会的约束，以及高等生物尤其是恒温动物生活条件的局限，使得对该课题的研究很大程度只停留在统计学的水平上，而很难深入到基因这一最根本的层次。

前面提过，鱼类广泛分布于世界上各个角落，其生存条件之宽松是其他生物所无法比拟的，借助于河豚鱼这一模式生物，我们可获得鱼类与人对应的各种基因，再通过分析这些基因（尤其是可能会对环境因子作用起关键反应的基因）在鱼类中不同环境条件下（如温度、PH、盐浓度、各种化合物等）的功能与表达差异，从而找到环境因子致病的实质。

四、克隆致病基因方法的有力补充：

人类基因组计划初期达到的目标是四张图的完成，不管其中含有多少遗传标记或多少序列，它还只是一张简单的由 A、T、G、C 四种碱基罗列的一本“天书”，并没有直接告诉我们其所蕴藏的真正秘密。现在我们已经建立了不少“解读天书”——识别基因的手段。四张图的完成为我们识别基因铺就了坚实的基础，而随着一个个基因的克隆，又将进一步丰富这四张图的内容。

总结迄今为止我们分离一千多个人类基因的技术，基本可分为三类：定位克隆（positional cloning）、候选克隆（candidate cloning）和功能克隆（functional cloning）。

1. 功能克隆：

人类基因组计划开始之前,人类基因的分离基本上全是按照功能克隆的策略成功分离的。人类的性状都是与蛋白质功能有关的。功能克隆就是从蛋白质功能着手,循着蛋白质—mRNA—cDNA—基因这一基因表达的反路径进行的,最成功的例子就是镰刀型细胞贫血病基因的克隆。病人的典型症状提示其发病可能与红细胞的某一蛋白质有关,通过免疫电泳等技术发现了异常的血红蛋白,根据蛋白质的氨基酸序列设计兼并的寡核苷酸探针以杂交筛选人的有核红细胞系的cDNA文库而得到了 α -珠蛋白基因的cDNA,再比较正常人与发病者cDNA,从而确定此基因即为该病的致病基因。

功能克隆的策略还反映在利用蛋白质抗体筛选cDNA文库技术、比较mRNA克隆法、差减杂交、差异显示等许多克隆技术方面。

功能克隆的最大优点是不依赖与基因组的信息,从而避免了由于各种基因组信息的缺乏或相互干扰而导致的许多麻烦。虽然功能克隆还具有技术成熟、简便经济等优点,但是由于实际生活中,人类的各种生理、病理现象和临床症状并不能提示某种疾病或某种生理功能是与哪个具体的蛋白质的功能有关,该蛋白质的组织特异性也不尽了解,更无法解释不同组织甚至是同一组织内的蛋白质与mRNA的差别是正常差异还是特异性表达造成的,这一切都限制了功能克隆在今后大规模进行基因组后(post-genome)研究方面的应用。

2.定位克隆:

定位克隆是随着遗传图的建立而发展起来的基因克隆策略。它的基本路线是:先建立某一表现型(主要是致病基因)与基因组中某一特定位置之间的联系,根据这个联系和物理图的作用,我们就将经典遗传学信息转变为物理标记—各种遗传标记所代表的确定的基因组区域,建立该区域的克隆重叠群(如YAC、BAC contig等);从该重叠群中筛选可表达的基因,即建立该区域的转录图,最后通过比较转录图中的各个成分在正常人与病人之间的区别,确立疾病与基因的联系而找到疾病相关基因。

该策略的优点正是避免了功能克隆的缺点而紧紧抓住表现型与基因这最根本的两方面,避开了多发病机制、多因子参与、何种组织和何种蛋白质的变化与之有关等因素的“误导”。

近几年来对一些致病基因分离,尤其是多基因病的致病基因的分离工作都是此策略的具体体现,如糖尿病致病基因、高血压致病基因、多指(趾)基因等等。

① 家系的选择

目前建立表现型和基因组位置之间联系的唯一有效途径就是通过对家系的分析。一般认为家系越大越好,一个有40多个可提供信息—减数分裂事件的三代以上的家系基本上可提供“肯定连锁”的证据。如哮喘病相关基因、乳腺癌相关基因和单发型开角青光眼基因等的克隆无不与好的家系紧密相连。正如此,疾病

家系已成为一种新型资源,被各国争夺与保护。有人说:(家系)是此时此刻人类遗传学之所在(This is where human genetics is at the moment)。

一般三代以上的家系要比二代家系多提供 20% 以上的信息量,可较容易确定同号染色体上各位点的等位基因在父系和母系染色体的粗略相位。遗传标记(STRP、SNP、EST、STS 等)的大规模分离即遗传图的丰富、计算机软件的开发、大规模 PCR 操作仪等仪器的制造,为以全基因组扫描法寻找靶区域创造了极好的条件。

② 靶区域的片段重叠群构建:

通过 STR 进行物理定位后,便是靶区域的片段重叠群的构建。以 YAC 构建的 contig 和部分区域的 BAC 或 PI contig 可以向资源中心直接购买或通过自行杂交、FISH 定位后确立。各种 contig 需严格验证方可使用。

③ 筛选编码序列:

从庞大的 Contig 中搜寻很少的编码序列是定位克隆的关键步骤。经典方法是用以靶区域的 YAC 或 BAC 等为探针筛选 cDNA 文库,此法简便易行,但面临的是严重的假阳性问题。在此基础上发展起来的 cDNA 筛选法(cDNA selection)^[49, 50, 51]是目前使用较多的方法。其原理是将靶区域的基因组 DNA 固定在尼龙膜上,同时通过 PCR 扩增 cDNA 文库获得 cDNA 插入片段作为探针,在一定的条件下二者进行杂交,然后洗去非特异性结合的 cDNA,再洗脱下特异性结合的 cDNA,并对其 PCR 扩增,克隆扩增产物,最后进行克隆鉴定、测序。此法简单、快速、有效。但由于预杂交、洗脱等非数字化控制的操作步骤,很容易导致 cDNA 亚文库出现大量的假阳性和假阴性;同时由于 PCR 扩增的高度敏感性与其自身技术不足,PCR 会扩增假阳性片段及倾向于扩增较小的片段,带来大量的重复性劳动和财力浪费,且结果有限。另外,这两种方法尚需了解基因的组织特异性及发育阶段的表达以选择不同种类的 cDNA 文库(目前一般选取基因数目较多的胎脑 cDNA 文库),这也大大限制了这两种方法的应用。

目前不依赖于 cDNA 来克隆编码序列的技术主要有外显子捕获^[52, 53](Exon-trapping)。简单地说,外显子捕获是利用大多数真核基因的转录本发生剪接的事实,将靶区域的基因组 DNA 的转录本在受体细胞内剪接,删除外显子之间的序列,再经 RT-PCR 扩增得到 DNA 片段中可能存在的外显子。该法最严格的要求就是载体的构建:首先要有一个强大的真核基因启动子以获得高度表达,其次是附加 exon-intron 结构以协助捕获。

外显子捕获技术的缺陷主要表现在其所用的载体上:(1)对能捕获的外显子的限制性,只有那些中间的、完整的、具有 SDg 和 SA_g 位点并在剪接盒中位于顺式的外显子才能被捕获,而一般真核基因的中间外显子都较小,较难进一步分析和筛选全长 cDNA;(2)接受基因组片段的局限性,只有在其两侧有与 BamHI 兼

容的酶切位点的外显子才能被捕获,外显子中间如有酶切位点则不能被捕获;(3)没有选择标记,因此不能鉴别有无插入片段;(4)含有隐含剪接位点而导致较高比例的假阳性。

外显子捕获和 cDNA 筛选获得的 cDNA 和 exon 都首先进行来源基因组片段的重新定位 (FISH 定位),在排除其他污染之后即进行测序,通过电子杂交检查是否与已知基因有一定的同源性等,然后再进行下一步实验。

此外,我们尚可采取其他一些办法如人鼠杂种细胞^[54、55]中直接克隆人的转录序列,以及同源重组分析法^[56、57]、特征序列检查法等来分析靶区域的编码。这里结合 Zoo-blotting 法,即筛选不同进化层次物种间都存在的保守序列,阐述河豚鱼作为模式生物在定位克隆中的作用,以作为其他方法的补充。

其依据的原理是:(1)功能重要的序列在进化上一般保守;(2)河豚鱼基因组与人类基因组之间存在着较大程度的同线保守。

具体的技术路线是:(1)获取靶区域或其附近尽量多的遗传标记(表达序列等单一序列);(2)以上述遗传标记为探针,筛选河豚鱼 BAC 或 cosmid 文库,获得该靶区域在河豚鱼中对应的区域 DNA,并搭建 contig;(3)对验证可靠的河豚鱼基因组 DNA 进行测序(由于河豚鱼的基因密度一般为 7-8Kb 一个基因,这样对河豚鱼对应区域的 DNA 全部测序要较 cDNA 筛选法、外显子捕获所得众多假阳性而造成的财力浪费要低,且效率也相应提高),或从河豚鱼 Genebank 中直接获取信息,并对序列进行电子杂交与剪接分析,最终获取河豚鱼该基因组 DNA 中的所有编码序列;(4)通过电子杂交或以上述获得的编码序列为探针杂交筛选人的 cDNA 文库,获得靶区域的大部分或全部编码序列;(5)分析数据,进行家系或人群验证,最终获得疾病相关基因。

此法的优点在于既在很大程度上避开了组织依赖性又可直接得到靶区域大部分 cDNA,而不会遇到外显子捕获的载体构建、cDNA 筛选中得到大量假阳性及所获 cDNA 片段偏小等麻烦。该法涉及技术简单、成熟、时效快,主要为杂交(包括 FISH 和电子杂交)、测序,而少用复杂的 PCR、克隆技术。

此法的主要缺点是需要丰富的靶区域探针(尤其是保守性特强的探针)以弥补异种同源性差而杂交困难。当然四张图的极大丰富将有效解决这一问题。

这样,原本由于其他模式生物与人之间的同线保守很差而近于荒废的 Zoo-blotting 技术又由于新型模式动物—河豚鱼的出现(同线保守较强)而“焕发出新的生命”。

3. 候选克隆:

候选克隆又分为定位候选克隆、功能候选克隆和定位功能候选克隆。是限于人类基因组计划的充分发展,已定位的基因越来越多的情况下派生出来的一种新的途径。其原理是利用未知疾病基因的定位区域和功能区域内的已知的一些编码

序列, 直接进行致病基因的筛选, 同时将功能已知但未与疾病建立联系的基因定位后, 再在定位区域搜索已在该区域定位的疾病, 决定二者是否有联系, 从而获得致病基因这一方法也归入此类。

其典型的例子就是美国 UCLA 的一个研究小组将早已发现的 Kinase 基因通过 FISH 的方法将其定位于 Xq21, 再从 Genebank 中查到 agamma-globinemia 已被定位于该区域, 通过对该基因在病人和正常人中的筛选, 最终确定了 Kinase 即是此病的致病基因。

没有任何一种克隆致病基因的方法可以代替其他所有的方法, 目前还必须综合利用各种方法才能更有效分离基因, 河豚鱼的加入无疑是起到了“添砖加瓦”的作用。

河豚鱼作为一低等的脊椎动物, 处于无脊椎生物和脊椎动物的交叉点, 对其作为模式生物后所进行的一系列基因分析, 将会给我们对基因进化、基因组进化甚至是生物进化的研究提供丰富的素材。

总之, 选择河豚鱼作为新兴模式生物并对其基因组进行系统研究将会极大地推动人类基因组计划和基因组后研究的进程。

参考文献

1. 杨焕明 21世纪医学生物学的蓝图—人类基因组计划, 科技导报 1996;9:9-18.
2. Olsen,M.V. et al (1986) Proc.Natl.Acad.Sci.USA 83: 7826-7830.
3. Sulston,J.et al (1992) Nature 356, 37-41.
4. Coulson,A.,Sulson,J., Brenner,S. et al ,(1986) Proc.Natl.Acad.Sci.USA83,7821-7825.
5. Oliver,S.G. et al Nature 357, 38-46 (1992).
6. 任淑仙 无脊椎动物学 1990, P7.
8. Kluge,A.G.et al 1977. Chordate structure and function. 2nd.ed Macmillian, New York.
9. Dennis,A. Powers Evolutionary genetics of Fish. Advances in Genomics vol 29.
9. Hinegardner,R. (1976) Evolution of genome size. In "Molecular Evolution"(F.Ayala, ed),pp.179-199. Sinauer,Sunderland,Massadrusetts.
10. Arkhipchuk,V.V. (1988) Effect of generation time on karyotype evolution in fish. *Gidrobiol.* 2h.24,92-94.
11. Ohno,S.(1974). Protochordata, cyclostomata and pisces,In "Animal cytogenetics. Vol 4:chordata 1"pp 1-91. Borntraeger, Berlin.
12. Bernadi,G., (1989) The isochore organization of the human genome. *Annu.Rev.Genet.*23:637-661.
13. Gardiner,K., Watkins,P., etal, (1988) Partial physical map of human chromosome 21 *Somatic Cell Mol. Genet.* 14:623-638.
14. Morizot,D.C.,(1990) Use of fish gene maps to predict ancestral vetebrate genome organization. In "Isozymes: structure , function, and use in Biology and medicine",pp 207-234, Wiley-Liss New York.
15. Hochachka,P.W., and Somero,G.N., (1984) Temperature adaptation. In "Biochemicl adaptaton " pp 355-449. Princeton Univ.Press. Princeton, New Jersey.
16. Hinegardner,R., etal, (1968) Evolution of Celluar DNA content in Teleost Fisher *Am.Natl.*102:517-528.
17. Pizon,V., etal(1984), Nucleotide sequence organization in the very small genome of a Tetraodontide fish,Arothron diadematus. *Eur.J.Biochem*,140:25-30.
18. 杨安峰 (1983): 脊椎动物学 p 146. 北京大学出版社.
19. Kadoo Miyaki, Osame Tabeta, Hiroshi Kayano(1995). Karyotype in six species of

- Pufferfish Genus Takifugu . Fisheries Science 61(4):594-598,
- 20 S.Brenner, G. Elgar, R.Sandford etal(1993), Characterization of the pufferfish(Fugu) genome as a compact model vertebrate genome. Nature 366:265-268.
 - 21 Chuoll-Honki(1984), Pufferfish available in Japan.
 - 22 Ben F.Koop, Joseph H.Nodeav(1996), Pufferfish and a new paradigm for comparative genome analysis. Pro. Natl. Acad. Sci. USA 93: 1363-1365.
 - 23 Maheshwar,M.M., Sandford,R., Nellist,M., etal (1996), comparative-analysis of the tuberous Sclerosis-2(TSC2)gene in human and puffefish. Hum.Mol.Genet. 5:133-137.
 - 24 Sandford,R., Sgotto,B., Aparicio,S., etal(1997), Comparative analysis of the polycystic kidney disease 1(PKD1) gene reveals an integral membrane glycoprotein with multiple evolutionary conserved domains. Hum.Mol.Genet.Sep:6(9):1483-1489.
 - 25 Elgar,G., Rattray,F., Greystrong,J., etal(1995), Genomic dtructure and nucleotide Sequence of the p55 gene of the puffer fish Fugu rebripes. Genomics 27:442-446.
 - 26 Baxendate,S., Abdlla.S., Elgar,G., etal(1995), Comparative sequence analysis of the human and pufferfish Huntingon's disease genes. Nature Genet. 10:67-76.
 - 27 Lim EH, Brenner,S., (1997), Short-range linkage relationships of the valyl-tRNA synthetase gene in Fugu rubripes. Immunogenetics 46(4):332-336.
 - 28 Buckler,A.J., Chang,D.D., Grae,S.L., etal(1991), Exon amplication: A strategy to isolate mammalian genes based on RNA splicing. Pro.Natl.Acad.Sci USA88:4005-4009.
 - 29 Yeo,G.S., Elgar,G., Sandford,R., etal(1997), Cloning and sequencing of complement C9 And its linkage to DOS-2 in the pufferfish Fugu rubripes.Gene Oct 24:203-211.
 - 30 Mason,P.J., Stevens,D.J., Luzzatto,L., etal(1995), Genomic structure and sequence of the Fugu rubripes glucose-6-phosphate dehydrogenase gene(G6PD). Genomics Apr 10;26(3):587-591 .
 - 31 Koh,C.G., Oon,S.H., Brenner,S., etal(1997), Serine/thereonine phosphatase of the Pufferfish Fugu rubripes. Gene. Oct 1;198(1-2):223-228.
 32. Greg Elgar, Richard Sanford, etal, (1996), Small is beautiful: comparative genomics With the pufferfish(Fugu rubripes). TIG vol.12 No.4:145-150.
 33. Senapathy,P., etal(1990), Splice junctions,branch point sites and exons:sequence statistics Identification and application to genome project. Methods Enzymol 183:252-278.
 - 34 Greg Elgar(1996), Quality not quantity: the pufferfish genome . Human Molecular Genetics Vol .5 review 1437-1442.

- 35 Trower,M.K.,etal(1996) Conservation of synteny between the genome of the pufferfish(*Fugu rubripes*) and the region on human chromosome 14(14q24.3) associated with familial Alzheimer disease(AD3 locus).*Pro. Natl. Acad. Sci.USA*93:1366-1369.
- 36 Copeland,N., etal(1993) *Science* 262:57-66.
- 37 Barendse, W.,etal(1994), *Nat.Genet.*6:227-235.
- 38 Rettenberger,G., etal(1995) *Genomics* 26:372-373.
- 39 Burt,D.W., etal(1995), *Trends Genet.* 11:190-194.
- 40 Nadeau,J.H., and Taylor,B.A.,(1984) *Pro. Natl. Acad. Sci.USA*81:814-818.
- 41 Sandford R., Sgotto,B., Burn, T., etal(1996), The tuberin(TSC2), autosomal dominant polycystic kidney disease (PKD1), and somatostatin type V receptor (SSTR5) genes form a synteny group in the *Fugu rubripes*. *Genomics* Nov 15;38(1):84-86.
- 42 Aparicio,S., Hawker,K., Cottage,A., etal(1997), Organization of the *Fugu rubripes* Hox clusters: evidence for continuing evolution of vertebrate Hox complexes. *Nat. Genet* May: 16(1):79-83.
43. Schofield,J.P., Elgar,G., Greystrom,J., etal(1997), Region of human chromosome 2(2q32-35)and mouse chromosome 1 show synteny with the pufferfish genome(*Fugu rubripes*). *Genomics* Oct 1;45(1):158-167.
44. Venkatesh,B., Brenner,S., etal(1995), Structure and organization of the isotocin and vasotocin genes from teleosts. *Adv.Exp.Med.Biol* 395:629-638.
45. How,G.F., Venkatesh,B., Brenner,S.,(1996). Conserved linkage between the puffer fish (*Fugu rubripes*) and human genes for platelet-derived growth factor receptor and macrophage colony-stimulating factor receptor. *Genome Res.* Dec;6(12):1185-1191.
46. Aparicio,S., Morrison,A., Gould,A., etal(1995), Detecting conserved regulatory elements with the model genome of the Japanese puffer fish, *Fugu rubripes*. *Proc.Natl. Acad,Sci USA* 92: 1684-1688.
47. Maheshwar,M.M., Standford,R.,etal(1996), Comparative analysis and genomic structure of The tuberous sclerosis 2(TSC2) gene in human and pufferfish. *Human Molecular Genetics* Vol.5,No.1 131-137.
48. Venkatesh,B., Si-Hoe,S.L., Murphy,D. etal(1997), Transgenic rats reveal functional conservation of regulatory controls between the fugu isolation and rat oxytocin genes. *Pro. Natl. Acad. Sci.USA*87: 12462-12466.
- 49 Parimoo,S.,etal,(1991), *Pro. Natl. Acad. Sci.USA*88:9623.
- 50 Lovett,M., etal(1991), *Pro. Natl. Acad. Sci.USA*88:9628.

- 51 Morgen, J.G., et al (1992), Nucl. Acid. Res. 20:3173.
- 52 Geoffreg M, Duyk, Surrón Kim, Richard M. Myers, et al (1990), Exon trapping: A genetic screen to identify candidate transcribed sequence in cloned mammalian genomic DNA. Pro. Natl. Acad. Sci. USA 87:8995-8999.
- 53 Alan J. Buckler, David D. Chang, Sharon L. Graw, et al (1991), Exon amplification: A strategy to isolation mammalian genes based on RNA splicing. Pro. Natl. Acad. Sci. USA 88:4005-9.
54. Laura Corbo, Julic A. Maley, et al (1990), Direct cloning of human transcripts with HnRNA from hybrid cell lines. Science 249:652-655.
55. Pu Liu, Randy Legerski, Michael J. Siciliano (1989), Isolation of human transcribed sequence from human rodent somatic cell hybrids. Science 246:81-85.
- 56 David M. Kurmit and Brian Sead (1990). Improved genetics selection for screening bacteriophage Libraries by homologous recombination in vivo. Pro. Natl. Acad. Sci. USA 87:3166-3169.
- 57 Roni J. Bollag, Alan S. Waldman R. Michael Liskey (1989), Homologous recombination in Mammalian cells. Annual review of genetics. 23:199-225.

综述二

人类多药耐药基因和多药耐药机制

人类基因组计划的前身—肿瘤计划的最主要工作就是肿瘤的治疗,在总结肿瘤计划的得失时,1986年由Dulbecco^[1]提出了人类基因组计划。肿瘤的治疗一般有手术治疗、放疗、化疗、和基因治疗,四种方法相辅相成,其中化疗是许多肿瘤目前唯一可行的治疗方法,如绒癌等。在化疗的过程中,人们发现(临床和实验室)肿瘤细胞能通过一种依赖能量的药物排出泵降低细胞内的药物浓度^[2]而对许多化疗药物产生耐药。这种现象首先于60年代由Lessel^[3]等人发现。这种多药转运子是由人的1型多药耐药基因(multi-drug resistance gene,MDR1)编码的位于细胞膜上的一种糖蛋白(P-glycoprotein)^[4]。这种糖蛋白的表达水平与药物的耐受水平及细胞内的药物浓度紧密相关。Gros等人通过将MDR1基因的全长cDNA转染到药物敏感的小鼠细胞株中,获得了多药耐药的表型(phenotype)从而建立了MDR1基因与多药耐药之间的联系,其他的一些实验也验证了多药耐药细胞系都有一共同的特征—P-glycoprotein的表达。肿瘤专家承认药物的耐受性是肿瘤治疗一尚未解决的重要问题。药物耐受性既表现在化疗一开始就存在的内源性耐药,又体现在化疗过程中出现的获得性耐药。目前,临床实践中已采取了许多办法来降低化疗药物的耐受性,诸如多种细胞毒药物的“鸡尾酒疗法”。但是最终要从根本上“攻克”耐药性这一难题,还必须从多药耐药基因及相关基因着手,从基因的结构、功能和表达、调控的研究着手,建立各种多药耐药性细胞系;寻找针对不同肿瘤的抗耐药药物—逆转耐药;研究、制备多药耐药糖蛋白的单克隆抗体;同时开展转基因治疗的研究与应用。

肿瘤细胞的耐药表型可随着诱导药物的不同、肿瘤细胞的种类、分化阶段的不同而表现出不同的耐药表型,可以是以某一耐药基因表达,也可以是多种耐药基因同时表达的多种耐药表型。

人类多药耐药基因

人的多药耐药基因家族有2个:MDR1和MDR3(有时又称为MDR2)。二者均位于7号染色体的7q21.1^[5,6],以头尾相接的方式连接,基因组跨度为230Kb,基因间序列约为34Kb,MDR1基因约为120Kb左右,MDR3为74Kb^[7],它们之间的同源性为85%。

MDR1 基因及其蛋白的结构

MDR1 基因全长约为 120Kb, 含有 29 个外显子, 其中 27 个编码 P-glycoprotein, 14 个编码 P-glycoprotein 的左侧序列, 13 个编码右侧序列^[9]; 28 个内含子, 其中 26 个位于基因编码区。基因全长 cDNA 约为 4.5Kb^[9]。各外显子和内含子的大小及其剪接位点见表一^[10]。

表一：外显子和内含子的大小及其剪接位点

exon/intron No.	Exon size(bp)	Exon end	Intron length	Intron type
- 1	>95	-330	>18000	
1	323	-7	500	
2	74	68	4300	2
3	49	117	1000	0
4	169	286	>20000	1
5	52	338	3100	2
6	192	530	541	2
7	172	702	4700	0
8	125	827	7500	2
9	172	999	2800	0
10	114	1113	250	0
11	111	1224	170	0
12	126	1350	200	0
13	204	1554	300	0
14	171	1725	3200	0
15	162	1887	850	0
16	177	2064	600	0
17	147	2211	2700	0
18	108	2319	1900	0
19	78	2397	2700	0
20	84	2481	4800	0
21	204	2685	1100	0
22	101	2786	1200	2
23	141	2927	2500	2
24	157	3084	1100	0

25	198	3282	5400	0
26	207	3489	3300	0
27	147	3636	1400	0
28	587	4223		

MDR1 基因有二个启动子（上游启动子和下游启动子）^[11]。上游启动子位于起始位点上游 70Kb^[6]，大多数正常组织的转录是从下游启动子开始的，下游启动子位于外显子 1 中，转录起始位点于起始密码子上游 140 核苷酸，位于转录起始点 5'端的外显子部分称之为 exon1a，3'端的称之为 exon1b，转录起始密码子 ATG 位于 exon2 中。转录位点有一 CAAT box^[12]，但缺少象大多数真核细胞一样的 TATA box^[13]；位点下游有两个 GC 富含区可识别转录因子 SP-1，位点上游有一个识别 AP-1 的回文区，具有增强子的功能；在 3'末端终止密码子 TGA 的下游有加尾信号结构。

MDR1 基因编码区的外显子大小范围为 49-587bp（也就是 16-69 个密码子），其平均大小为 47.5 个密码子，与其他基因的外显子平均值 44.5 密码子基本相符^[14]。所有的剪接位点都遵循 GT/AG 法则^[15]。

内含子按其位置分为三类：位于一密码子之后的为 0 型；位于一密码子第一核苷酸之后的为 1 型；位于一密码子第二核苷酸之后的为 2 型。编码区的 26 个内含子中 19 个为 0 型，1 个为 1 型，6 个为 2 型。它们在整个基因的分布并无规律，其中编码 N 端半侧膜结合区的 8 个内含子中，4 个 2 型、3 个 0 型、1 个 1 型；编码 C 端半侧膜结合区的 8 个内含子中，6 个 0 型、2 个 2 型。

MDR1 编码的蛋白质 P-glycoprotein 由 1280 个氨基酸组成，N 端未甲基化时分子量为 140Kd^[16]，甲基化后为 170Kd。P-glycoprotein 有 12 个跨膜结构和二对核苷酸结合位点(nucleotide-binding site)。1—637 氨基酸组成了 P-glycoprotein 的 N 端半侧；638—1280 氨基酸构成了 C 端半侧；N 端和 C 端各有 6 个跨膜结构和一对核苷酸结合位点（图 1）^[17]。

N 端半侧和 C 端半侧对称比较发现（图 2）^[10]43%的氨基酸（263/613）是一致的，35%（215/613）的氨基酸虽然结构不一样，但功能是一样的；在 DNA 水平上，其碱基同源性的为 52%。二对核苷酸结合位点，即 N 端的 529-591 氨基酸和 419-446 氨基酸分别与 C 端的 1174-1236 氨基酸和 1061-1088 氨基酸高度保守一致，同源性分别为 90.3%(56/62)和 77.8%(21/27)。N 端半侧的 1-48 氨基酸为亲水区组成 N 端，49-350 氨基酸为疏水区组成了 6 个跨膜结构，351-637 氨基酸为亲水区组成了 N 端半侧的核苷酸结合位点，638-708 氨基酸为 N 端半侧与 C 端半侧的连接区；709-993 氨基酸为疏水区组成了 C 端半侧的 6 次跨膜结构，994-1280 为亲水区组成了 C 端；氨基酸 91-93、94-96、99-102 组成了三个糖基化区域。

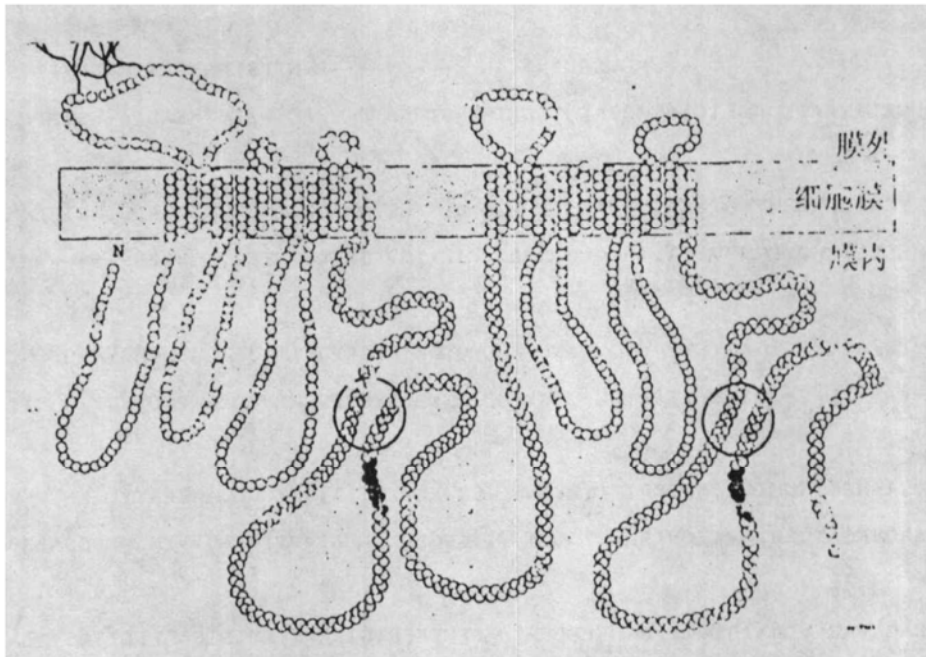


图 1: P-glycoprotein 的结构示意图。

细黑线表示糖基化点；圆圈表示 ATP 结合位置；黑点表示 ABC 家族特征区。

MDR3 基因及其转录特性⁷⁾

MDR3 基因为 74Kb, 含有 28 个外显子, 其中 27 个有编码功能, 大小范围 55—259bp, 平均为 145bp。第一外显子中含有一富含 GC 的非翻译序列, 有证据表明存在 exon-1。起始密码子位于 exon2 中, 编码的蛋白质与 MDR1 一样含有 12 个跨膜结构和二对核苷酸结合位点, N 端半侧和 C 端半侧互相对称。N 端半侧的 3 和 4 跨膜结构由同一外显子编码, C 端半侧的 9 和 10 跨膜结构亦由同一外显子编码。其余的 8 个跨膜结构由不同的外显子编码; 外显子 12、13、14 和 25、26、27 分别编码 N 端半侧和 C 端半侧的核苷酸结合位点。27 个内含子中含 20 个 0 型、1 个 1 型和 6 个 2 型, 其中只有 9 号和 23 号内含子以及 13 号和 26 号内含子具有相同的位置。

MDR3 基因的剪接有四种方式, 均发生于 exon23 和 exon26 中 (图 3)。它的发生可能与内含子和外显子中隐含的剪接受体位点有关, 初步研究认为组织中的 MDR3 以 A 型和 C 型居多。由于剪接方式的改变 (甚至导致编码 11 号跨膜结构的外显子 23 的整个丢失) 造成 MDR3 蛋白的拓扑结构改变, 这对于其功能的研究将具有重要意义。

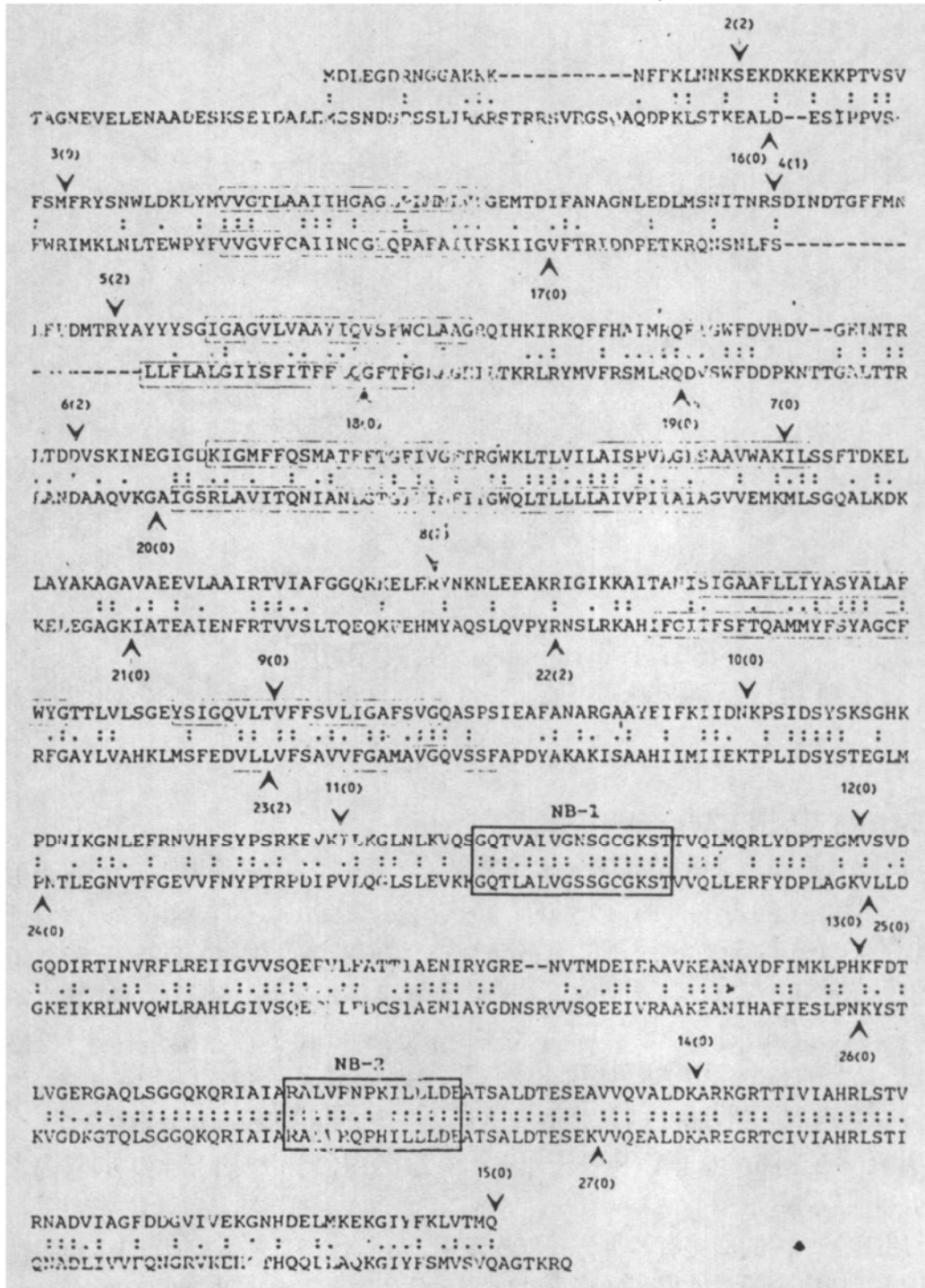


图2: N端半侧和C端半侧的氨基酸同源性比较

箭头代表内含子的位置和类型；细线框代表跨膜结构区；粗线框代表核苷酸结合位点（NB-1和NB-2）；双点代表同样的核苷酸；单点代表功能相近的核苷酸

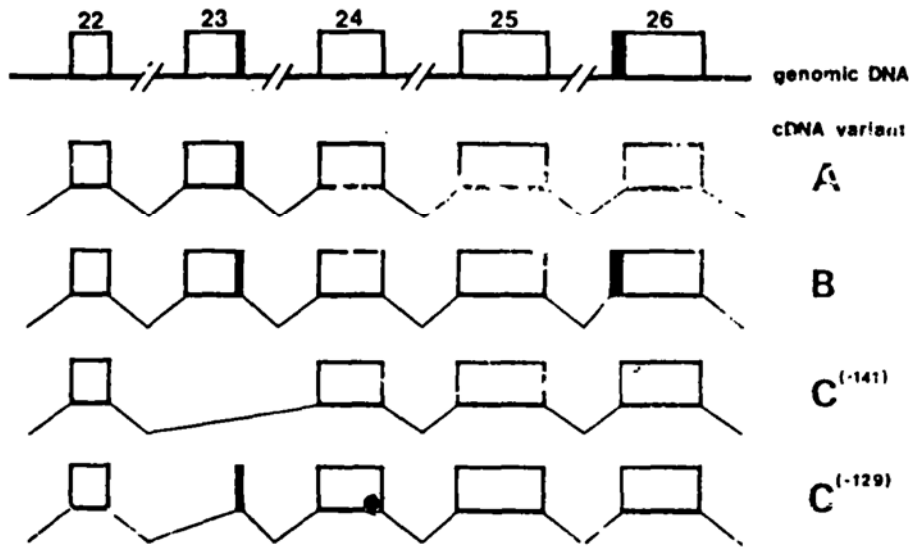


图 3: MDR3 的四种转录方式

通过对人的 MDR1 和 MDR3 基因的序列比较,二者从外显子 6 至外显子 27 都具有完全相同的碱基长度和位置。Exon28 在 MDR1 中为 587 个碱基,在 MDR3 中为 259 个碱基,这可能是 3'非翻译区长度不一致造成的。各内含子的大小在 MDR1 和 MDR3 中未见有保守性,尤其是内含子 4 在 MDR1 中>20Kb,而在 MDR3 中为 8Kb。

MDR3 在人体中表达的组织很有限,仅在肝脏等少数组织中可检测到。至今我们仍未发现 MDR3 参与多药耐药作用的证据。

MDR1 基因和 MDR3 基因的进化

由于 MDR1 和 MDR3 基因及其相应的蛋白质结构有着明显的两侧(3'和 5', N 端和 C 端)重复性,人们设定了一原始基因(primordial gene)^[9,18,19]内复制模式(internal duplication)来解释这种结构。1982 年 Higgins^[20]、Gilson^[21]等人发现细菌蛋白 hisP、malK、OppD^[22]、pstB^[23]都是膜蛋白,都是依靠 ATP 提供能量而负责将胞浆内的一些物质运输到胞浆外。这些细菌膜蛋白的结构高度相似,都有 6 个跨膜结构和一对核苷酸结合位点,其中 hisP 与 MDR1 有 33%的氨基酸一致,另有 28%的氨基酸功能一致,尤其是核苷酸结合位点的氨基酸在各种细菌蛋白和人的 P-glycoprotein 之间高度同源。鉴于上述特点,可将这组结构称之为 ABC 超级家族(ATP-binding cassette superfamily)(表二)^[24]。

表二: ABC 超级家族成员

细菌营养转运子:	malk(maltose); hisP(histidine)
肽/药物转运子:	hylB(hemolysin); OppF 和 OppD(寡肽); MDR1; pfmdr(镰形疟原虫); STE6(yeast sex peptide alpha) Tap1 和 Tap2(antigen presentation)
未知功能:	CFTR(cystic fibrosis); MDP(multi-drug resistant associated protein)

这些现象给内复制理论提供了根据。一般认为内复制之后,复制子之间的内含子位置应是高度保守的^[25],但发现在 P-glycoprotein N 端半侧和 C 端半侧仅有二对内含子的位置一致(图 2)^[10]。人们设定了各种模型以弥补内复制理论在该问题上的不足。1 内含子的丢失:原始基因内复制后原有的一些内含子分别从 N 端半侧和 C 端半侧丢失。一般内含子的丢失应是丢失非同源部分而尽量保留同源部分,再者,内含子的丢失会造成大外显子的形成,但事实并不与此相符;2 外显子插入:这可以解释 exon5 很小且无同源序列,但不能解释 5'端和 3'端外显子之间的高度保守性;3 剪接滑动造成内含子位置改变,但这并不能解释内含子类型的差异等问题,因为剪接滑动会造成内含子位置的改变但一般都不会改变内含子的类型;4 内复制之后内含子的插入,但不能解释为什么 exon 平均大小不变及不同类型的内含子并非随机分布的问题。鉴于上述这些模型的种种缺陷,Chang-jie Chen^[10]等人设计了另一全新模型:两侧的核苷酸结合位点来自于同一个或相似的原始基因;两侧的跨膜结构来自于不同的原始跨膜结构基因,每个原始核苷酸结合位点基因与原始跨膜结构基因融合为单侧结构,这个单侧结构的原始状态即为上述细菌转运蛋白,两个单侧结构融合后形成 MDR1 基因(图 4)。

N 端半侧的右端可认为是含有不同内含子类型(0 型和 2 型)的外显子的融合,而左端可认为是含 2 型内含子的外显子融合后,插入了一新的内含子或新的剪接位点形成,左端和右端融合后形成 N 端半侧的跨膜结构,当然左端和右端的序列同源性可用其他的理论来协同加以解释。上述模型似乎可用来完整地解释 MDR1 基因和 MDR3 基因的结构。该模型支持 N 端半侧和 C 端半侧是相互独立工作的假设,似乎为 P-glycoprotein 可转运各种物质找到了一个较好的解释理由。但是,某些实验证明无论是哪一侧的核苷酸结合位点的碱基突变后,整个基因的转运功能就会丧失^[26],这说明 N 端半侧和 C 端半侧并不是独立工作的,这似乎又是对上述模型的“沉重打击”。

看来,单纯用某一种机制来解释 MDR 基因的进化过程似乎都不够全面,MDR 基因可能是多种进化机制协同作用后形成的。

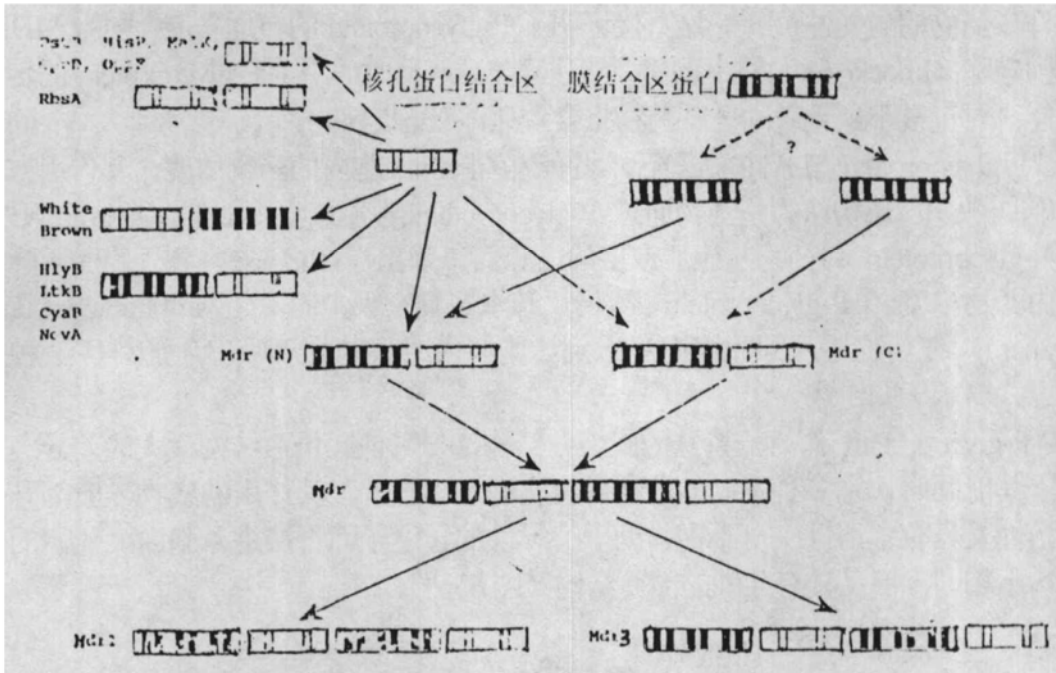


图4: MDR1 进化模型

黑线代表膜结合区基因; 虚线代表核苷酸结合位点基因

尽管这样, 努力去阐明该基因的结构和进化将有助于我们最终解释清楚为什么 P-glycoprotein 可以作用于大量的互不相干的物质。

另外, 多药耐药基因在其他生物中基本都存在, 如小鼠和中国仓鼠中有 3 种亚型^[18] (mdr1、mdr2、mdr3)。在进化上我们可将 ABC 家族基因称为“看家基因” (keeping-house gene)。

P-glycoprotein 的功能

MDR1 基因的编码产物为 P-glycoprotein, 该蛋白具有下列特征: 1 依靠 ATP 提供能量才能具有活性; 2 降低细胞内的药物浓度; 3 作用底物多种多样。

Dano^[27]等人发现当体外培养的耐药细胞株中的 ATP 用尽后, 或者将培养基中去除糖份, 或者在培养基中加入能量代谢抑制剂后, 耐药细胞株将表现出不耐药特征, 这表明 P-glycoprotein 的作用需要依靠 ATP 提供能量才可进行^[28,29]。利用放射性标记的 ATP 底物, 将会特异性标记 P-glycoprotein, 而放射性标记的 ADP 等类似底物却不能特异标记 P-glycoprotein, 也就是说不能与 P-glycoprotein 结合, 这直接证明了 P-glycoprotein 与 ATP 之间的密切关系。MDR1 本身特异性的结构也说明了这一特征。

将 MDR1 基因的全长 cDNA 转移到药物敏感细胞株中将使得该细胞株获得

对许多药物的耐受性^[30,31,32]，这直接证明了 P-glycoprotein 的转运功能；同时 MDR1 基因敲除 (knock out) 的小鼠模型的建立^[33]又从反面证明了 P-glycoprotein 的转运功能，对正常小鼠无效的一些药物将会致死该 knock-out 小鼠。

P-glycoprotein 与 ATP 结合后，将有效的降低细胞内的药物浓度，其作用方式可能是^[34]：1. 最初认为是药物通过 P-glycoprotein 在胞膜上形成的单孔排到胞外；2. P-glycoprotein 将药物从胞质双层的内膜面“弹出” (flippase) 到外膜面而降低胞内药物浓度；3. P-glycoprotein 类似于“疏水区真空吸尘器” (hydrophobic vacuum cleaner)，在药物透入细胞膜而尚未到达胞浆内时将药物“吸出”胞外。三种作用方式简示于图 5^[24]。

P-glycoprotein 作为一特殊的膜蛋白结构，与其他膜蛋白结构最大的差异之处在于其能同时识别多种不同的药物并将它们泵出胞外。其作用的底物一般都既有脂溶性又具有水溶性，可不需借助于特异性的转运子而直接进入胞内。他们的结构各不相同，但又具有共同的闭环结构 (图 6)^[24]。

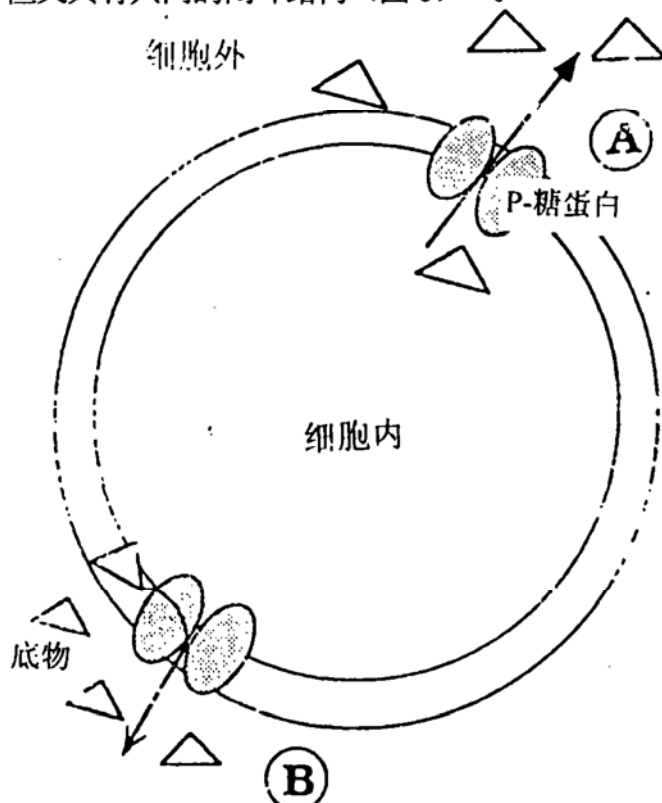


图 5: P-glycoprotein 排出药物的方式示意图 (A 代表方式 1 和 2; B 代表方式 3)

长期以来，对于不同的药物如何与 P-glycoprotein 作用这一问题一直困扰着科学家们。最近的药物标记实验证明 P-glycoprotein 与药物的结合部位分布于整个分

子上,不同的药物可能结合于不同的区域^[35],关键的结合区域主要位于跨膜结构3、5、6、11和12的胞浆及跨膜部分^[36],其他的一突变实验表明核苷酸结合位点附近的区域^[37]甚至是细胞外表面的蛋白也参与底物结合形成^[38,39]。

另外,有实验证明 P-glycoprotein 尚有氯化物通道活性,该通道活性可与转运活性相互转换^[40]。

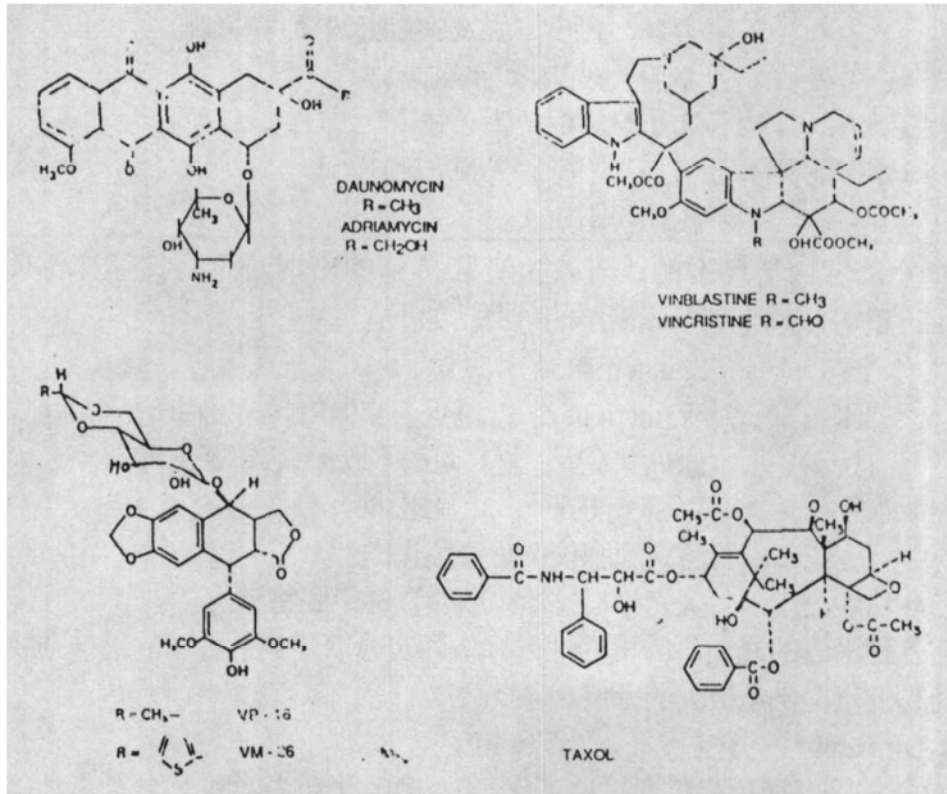


图6: P-glycoprotein 部分底物化学结构

P-glycoprotein 的表达

通过免疫组化的方法发现在人体的各组织中,MDR1 的表达各不相同(表三)。

在各种组织、器官中的表达水平的不同也揭示出 P-glycoprotein 的不同作用;

- 1 P-glycoprotein 在肝胆管表面和肾近曲小管、及肠道等处表面呈极性分布^[35,48,49],这说明其可能与排泄人体自身的脂溶性代谢产物和外源性抗生素有关^[50];
- 2 造血干细胞的表达可能与降低骨髓抑制有关^[51];
- 3 在脑、睾丸、妊娠子宫等处高度表达,可能与其作为血-脑、血-睾丸、母

体-胎儿的屏障而免受一些外源抗生素或有毒的内源性代谢产物的影响有关^[52,53,54];

4 与甾醇、孕酮直接作用^[55]。

表三: MDR1 在正常组织、器官中的表达水平^[241,42,43,44,45,46,47]

高度表达组织、器官:	肾上腺皮质、肾近曲小管、肝胆管表面、胎盘、结肠、小肠、人脑皮质细胞、睾丸内皮细胞、胰腺上皮细胞、巨噬细胞;
中度表达组织、器官:	肾上腺髓质、气管、肺(主支气管)、前列腺;
低度表达组织、器官:	皮肤、骨骼肌、心脏、脾、食管、胃、卵巢、脊髓、骨髓

随着分子生物学技术的改进及单克隆抗体的出现, 现有检测 MDR1/ P-glycoprotein 的方法越来越多, 简单的可概括如下:

DNA 水平:	Southern blot ^[56,57] ;
RNA 水平:	Northern blot ^[58] ; 斑点杂交; RNA 酶保护法 ^[59] ; 原位杂交 ^[60] ; 逆转录酶多聚酶链式反应 ^[61] ;
蛋白质水平:	Western blot ^[62] ; 免疫组化 ^[63] ; 流式细胞仪; 免疫荧光;
蛋白质功能测定:	荧光染料排泄; 放射性标记底物转运。

上述各种手段的建立, 为更好地进行体外实验以寻找抗肿瘤的最佳药物搭配和最佳药物剂量提供了极大的方便。

P-glycoprotein 的调控

MDR1 基因翻译后尚需通过糖基化^[64,65]、磷酸化和去磷酸化^[66,67,68]后才能获得有效活性。

实验发现许多潜在的 MDR1 基因转录调控因子, 包括许多相似的热休克因子^[69]等等。热休克、CdCl₂、砷盐能诱导 MDR1 转录。通过 MDR-CAT 报告基因发现突变的 Ras 和 P53 基因可刺激 MDR1 启动子^[70], 一些研究表明包括 C-Raf 激酶的信号传导系统可调控 MDR1 基因的转录^[71]。

利用二甲亚砜等诱导剂处理人结肠癌细胞系可增加 MDR1 基因的表达^[72]。一些细胞毒因子、紫外线或部分肝切除都能在人的细胞株中增加 MDR1 基因的表达^[73,74,75]。一些药物介导的诱导 MDR1 表达行为可被非特异性蛋白激酶抑制剂阻断。

在研究摸索化疗方案的过程中发现一些药物可以逆转 MDR1 基因的多药耐药特性(表四)^[51]。这种“负调控”的机制并不是通过调控 MDR1 基因的表达, 而是可能通过与 P-glycoprotein 结合后产生的竞争抑制作用降低了 P-glycoprotein 的

多药耐药性。

多药耐药机制

人体各种药物的耐药机制并不完全是由 P-glycoprotein 来实现的, 其他的一些机制也参与了多药耐药: 1. 药物在亚细胞水平的分布可改变靶器官的药物浓度; 2. 药物靶器官的转换; 3. DNA 修复能力的不同; 4. 与其他因子协同解毒的细胞代谢系统的改变。这些机制可相互作用或单独作用, 但都会降低药物浓度而产生多药耐药的表型。

表四: 逆转 P-glycoprotein 耐药特性的部分药物

类固醇激素:	可的松、醛固酮、孕酮、地塞米松等
钙拮抗剂:	维拉帕米、硝苯吡啶、硫氮酮、尼卡地平
心血管药物:	奎尼丁、胺碘酮
免疫抑制剂:	环孢菌素 A 等
抗抑郁药:	硫哌啶酮
抗疟疾药:	奎宁、阿的平
钙调蛋白抑制剂:	氟奋乃静、三氟拉嗪等
抗生素:	红霉素、头孢三嗪、先锋必
抗真菌药:	酮康唑、雷帕霉素

Beck^[76]和 Danks^[77,78]等人发现在 CCRF-CEM 人 T 细胞白血病细胞株中, 虽然不表达 P-glycoprotein, 但一样存在多药耐药的表型, 研究显示该耐药机制可能与 DNA 拓扑异构酶 II 的改变有关。

Cole^[79]等人发现 MRP 可能在多药耐药机制方面起着重要作用。MRP 基因为 7.8—8.2Kb, 为位于 16p13.1, 其产物为 190Kd 的膜结合糖蛋白^[80]。

Morrow^[81]等人发现谷胱甘肽-S-转移酶可有效降低药物毒性而从侧面起到多药耐药的表型。

某些药物诱导的细胞凋亡程序的改变, 包括 Bcl-2 通道中基因的改变, 也会产生一些多药耐药的表型^[82]。

在其他实验中尚发现另外一些因素参与多药耐药^[83,84,85,86,87]。

总之, 多药耐药的机制是相当复杂的, 单靠 MDR1 并不能说明所有的问题, 多药耐药是一种综合因素作用的表型。

多药耐药基因的临床应用

通过对各种耐药机制的研究将为临床治疗各种肿瘤、各种传染性疾病的找到更

好的途径。对于大多数化疗药物来说,增加其剂量可提高肿瘤的治愈率,对于一些难治性肿瘤和非霍杰金淋巴瘤,强烈化疗能获得缓解和提高患者生存率。但由于化疗导致骨髓抑制使许多抗肿瘤药物的使用剂量有限而难以发挥其全部作用,严重的骨髓抑制可并发感染而不得不在化疗获得效果之前停止给药。骨髓抑制主要表现为外周血三系细胞的明显下降,尤其是白系,抑制可分为直接的和间接的,直接抑制是化疗药物直接作用于造血干细胞而影响其复制和分化,间接抑制是化疗药物作用于造血微环境而改变造血干细胞的复制和分化。已有的一些减轻骨髓抑制的办法如骨髓移植、单克隆抗体导向药物、脂质体包裹药物、区域性化疗、造血因子等等,都只能作为“姑息疗法”,并不能从根本上获得大剂量化疗的有效保障,也就很难获得肿瘤的治愈。鉴于MDR1基因的广谱耐药性,将其转移到造血干细胞后,可使造血干细胞获得对多种化疗药物的耐受性。将基因转移后的造血干细胞回输给病人,可进行反复大剂量化疗而不会出现骨髓抑制,从而提高肿瘤的治愈率和病人的长期生存率。

同时利用MDR1基因转移建立的细胞系和动物模型,将在药学的研究上具有重要的价值,利用MDR1基因的多药耐药特性使其作为构建某些特殊重组体时的特殊的选择标记。

检测肿瘤本身的多药耐药蛋白表达量的高低及在肿瘤治疗中多药耐药表型的改变将有助于预测病情和改变治疗方案。

参考文献

1. Dulbecco. Science 1986.
2. Ling,V., and Thompson,L. Reduced permeability in CHO cells as a mechanism of resistance to colchicine. J.Cell. Physiol. 83, 103-116 (1973).
3. Kessel,D., and Bosmann,H.B. On the characteristic of actinomycin D resistance in L5178Y Cells. Cancer Res. 30, 2695-2701(1970).
4. Juliano,R.L., and Ling,V. A surface glycoprotein modulating drug permeability in Chinese Hamster ovary cell mutants. Biochim. Biophys. Acta 455: 152-162 (1976).
5. Callen,D.F., Baker,E., Simmers,R.N.,et al. Location of the human multiple drug resistance Gene, MDR1 ,to 7q21.1. Hum. Genet. 77,142-144(1987).
6. Chin,J.E., Soffir,R., Noonan,K.E.,et al. Structure and expression of the human MDR(P-glycoprotein) gene family. Mol. Cell. Physiol. 152: 87-94(1992).
7. Carsten R., Lincke, Japp J.M. Smit et al. Structure of the human MDR3 gene and physical mapping of the human MDR locus. J. Biol. Chem. 1991, 266:5303-5310.
8. Roninson,I.B., Structure and evolution of P-glycoprotein. In "molecular and cellular biology of multidrug resistance in tumor cells", pp.189-211. Plenum, New York, 1991.
9. Chen,C.J., Chin,J.E., Ueda,K., etal. Internal duplication and homology with bacterial transport proteins in the mdr1(P-glycoprotein)gene from multidrug resistance human cells. Cell 47,381-389(1986).
10. Chang-jie Chen, Douglas Clark, Kazumitsu Ueda , et al. Genomic organization of the human multi-drug resistance (MDR1) gene and origin of P-glycoprotein. J. Biol. Chem. 1990, 265: 506-514.
11. Ueda,K., Pastan,I., and Gottesman,M.M., Isolation and sequence of the promoter region of the human multidrug resistance (P-glycoprotein)gene. J.Biol. Chem. 262, 17432-36(1987).
12. Ueda,k,Clark,D.P., Chen,C.J., etal. The human multidrug resistance (mdr1) gene. CDNA cloning and transcription initiation. J.Biol. Chem. 262: 505-508(1987).
13. Maniatis,T., Goodbourn,S., and Fischer,J.A., Regulation of inducible and tissue-specific gene expression. Science 236,1237-1245(1987).
14. Traut,T.W. (1988) Proc. Natl. Acad. Sci. USA85, 2944-2948.
15. Mount,S.M. (1982) Nucleic. Acids. Res. 10,459-472.
16. Kane,S.E., Pastan,I., and Gottesman,M.M., Genetic basis of multidrug resistance of tumor cells. J. Bioenerg.Biomembr. 22,593-618 (1990).

17. Balazs Sarkadi, Marianna Muller, Zsolt Hollo. The multidrug transporters—protein of an ancient immune system. *Immunology letter*, 54(1996): 215-219.
18. Gros,P., Croop,J., and Housman,D., Mammalian multidrug resistance gene: Complete cDNA sequence indicates strong homology to bacterial transport proteins. *Cell* 47, 371-380(1986).
19. Gerlach,J.H., Endicott,J.A., Juranka,P.F., etal , Homology between P-glycoprotein and a bacterial haemolysin transport protein suggests a model for multidrug resistance. *Nature* 324:485-489(1986).
20. Higgins,C.F., Haag,P.D., Nilaido, K., et al. Complete nucleotide sequence and identification of membrane component of the histidine transport operon of *S.Typhimurium*. *Nature* 298: 723-727.
21. Gilson.E., Nilaido,H., and Hofnung,M. (1982) Sequence of MalK gene in *E.coli.K12*. *Nucl.Acids.Res.*10, 7449-7458.
22. Higgins,C.F., Hiles,I.D., Whalley,K., and Jamieson,D.J. (1985) Nucleotide binding by membrane components of bacterial periplasmic binding protein-dependent transport system. *EMBO J.*4 1033-1040.
23. Surin,B.P., Rosenberg,H., and Cox,G.B.(1985) Phosphate-specific transport system of *E.coli* nucleotide sequence and gene-polypeptide relationships. *J. Bacterial*, 161:189-198.
24. Ira Pastan, and Michael M. Gottesman. Multidrug resistance . *Annu Rev. Med.* 1991,42:277-286.
25. Strathdee,C.A. and Lo,R. Y. (1989) *J.Bacterial.* 171,916-928.
26. Azzaria,M., Schurr,E., and Gros,P., Discrete mutations introduced in the predicted nucleotide-binding sites of the *mdr1* gene abolish its ability to confer multidrug resistance. *Mol.Cell. Biol.* 9,5289-5297(1989).
27. Dano,K., Active outward transport of daunomycin in resistant Ehrlich ascites tumor cells.*Biochim. Biophys. Acta* 323, 1466-1483(1973).
28. Cornell,M.M., Safa,A.R., Felsted,R.L., etal. Membrane vesicles from multidrug resistant human cancer cells contain a specific 150- to 170Kda protein detected by photoaffinity labeling. *Proc. Natl. Acad. Sci. USA*83: 3847-3850(1986).
29. Cornwell ,M.M., Tsuruo,T., Gottesman,M.M.,Pastan,I. 1987. ATP-binding of P-glycoprotein from multidrug resistant KB cells . *FASEB J.*1:51-54.
30. Gros, P., ben,N. Y., Croop,J.M., and Housman,D.E., Isolation and expression of a complementary DNA that confers multidrug resistant. *Nature*;323:78-731(1986).
31. Guild,B., Mulligan,R., Gros,P., etal, Retroviral transfer of a murine cDNA for

- multidrug resistant confers pleiotropic drug resistant without prior drug selection. Proc. Natl. Acad. Sci. USA85:1595-1599(1988).
32. Pastan,I., Gottesman,M.M., Ueda,K., et al, A retroviral carrying an MDR1 cDNA confers multidrug resistance and polarized expression of P- glycoprotein in MDCK cells. Proc. Natl. Acad. Sci. USA85:4486-4490(1988).
 33. Borst, P. and Schinkel, A.H. (1996) Eur. J. Cancer. 32A. 985-990.
 34. Gottesman,M.M., and Pastan,I., Biochemistry of multidrug resistant mediated by the multidrug transporter. Annu. Rev.Biochem. 62:385-427(1993).
 35. Germann,U.,(1996) Eur.J.Cancer 32A. 927-944.
 36. Muller,M., Bakos,E., Welker,E.,et al.(1996) J. Biol. Chem. 271:1877-1883.
 37. Hoof,T., Demmex, A., Hadam, M.R., et al. (1994) J. Biol. Chem. 269: 20575-20583.
 38. Zhang, X., Collins, K.I., and Greenberger, L.M.(1995) J. Biol. Chem. 270: 5441-5448.
 39. Welker,E., Sazabo,K., Hollo,Zs.et al (1995) Biochem. Biophys. Res. Comm. 216: 602-609.
 40. Gill,D.R., Hyde,S.C., Higgins,C.F., et al, Separation of drug transport and chloride channel function of the human multidrug resistance P-glycoprotein. Cell 71:23-32(1992).
 41. Vander, Valk,P., van Kalken,C.K., Kerelaars,H., et al, Distribution of multidrug resistant-associated P-glycoprotein in normal and neoplastic human tissues. Analysis with 3 monoclonal antibodies recognizing different epitopes of the P-glycoprotein molecule. Ann. Oncol. 1,56-64(1990).
 42. Chaudhary,P.M., and Roninson,I.B., Expression and activity of P-glycoprotein, a multidrug efflux pump in human hematopoietic stem cells. Cell 66,85-94(1991).
 43. Cardon-Cardo C, Obrien J, Bucci J.etal . Expression of the multidrug resistance gene product (P-glycoprotein) in human normal and tumor tissues. J. Biol. Chem 1990; 38: 1277-80.
 44. Lieberman DM, Reithmeier RAF, Ling V, et al. Identification of P-glycoprotein in renal brush border membranes. Biochem. Biophys. Res. Comm 1989; 162: 244-252.
 45. Nooter K, Herweijer H. MDR genes in human cancer. Br J Haematol 1992; 81:145-152.
 46. Marie,J.P., Brophy,N.A., Ehsan,M.N., et al, Expression of multidrug resistance gene mdr1 mRNA in a subset of normal bone marrow cells. Br.J.Haematol.81, 145-152(1992).
 47. Thiebaut,F., Tsuruo,T., Hamada,H., et al, Cellular location of the multidrug resistant

- gene product P-glycoprotein in normal human tissues. *Proc. Natl. Acad. Sci. USA* 84:7735-7738(1987).
48. Arias, I.M., Multidrug resistant genes, P-glycoprotein and the liver. *Hepatology* 12, 159-165(1990).
 49. Pastan, I., and Gottesman, M.M., Multidrug resistance in human cancer. *N. Engl. J. Med.* 316:1388-1393(1987).
 50. Rothenberg, M., and Ling, V., Multidrug resistance: Molecular biology and clinical relevance. *JNCL. J. Natl. Cancer. Inst.* 81:907-910(1989).
 51. Bert L. Lum, Pharm.D., George A., et al. Clinical trials of modulation of multidrug resistance *Cancer Supplement* December 1, 1993, Vol 72, No.11 3502-3514.
 52. Cordon, C.C., Obrien, J.P., Caslas, D., et al, Multidrug resistance gene(P-glycoprotein) is expressed by endothelial cells at blood-brain barrier sites. *Proc. Natl. Acad. Sci. USA* 86:695-698(1989).
 53. Tatsuta, T., Natio, M., Oh hara., et al, Functional involvement of P-glycoprotein in blood-brain barrier. *J. Biol. Chem.* 267:20383-20391(1992).
 54. Arceci, R.J., Croop, J.M., Horwitz, S.B., et al, The gene encoding multidrug resistance is induced and expressed at high levels during pregnancy in the secretory epithelium of the uterus. *Proc. Natl. Acad. Sci. USA* 85:4350-4354(1988).
 55. Qian, X.D., and Beck, W.T., Progesterone photoaffinity labels. P-glycoprotein in multidrug resistant human leukemic lymphoblasts. *J. Biol. Chem.* 265:18753-18756(1990).
 56. Ito, Y., Tanimoto, M., Kumazawa, T. et al, Increased P-glycoprotein expression and multidrug resistant gene(mdr1) amplification are infrequently found in fresh acute leukemia cells. Sequential analysis of 15 cases at initial presentation and relapsed stage. *Cancer* 63:1534-1538(1989).
 57. Merkel, D.E., Fuqua, S.A., Tandon, A.K., et al, Electrophoretic analysis of 248 clinical breast cancer specimens for P-glycoprotein overexpression or gene amplification. *J. Clin. Oncol* 7, 11, 1129-1136(1989).
 58. Fojo, A.T., Ueda, K., Salmon, D.J., et al, Expression of a multidrug resistance gene in human tumors and tissues. *Proc. Natl. Acad. Sci. USA* 84:2665-2669(1987).
 59. Goldstein, L.J., Galski, H., Fojo, A., et al, Expression of a multidrug resistance gene in human cancer. *JNCL. J. Natl. Cancer. Inst.* 81:116-124(1989).
 60. Shen, D. W., Pastan, I., and Gottesman, M.M., In situ hybridization analysis of acquisition and loss of the human multidrug-resistance gene. *Cancer Res.* 48:4334-4339(1988).

61. Futscher, B. W., Blake, L. L., Gerlach, J. H., et al, Quantitative polymerase chain reaction analysis of *mdr1* mRNA in multiple myeloma cell lines and clinical specimens. *Anal. Biochem.* 213:414-421(1993).
62. Bell, D. R., Gerlach, J. H., Kartner, N., et al, Detection of P-glycoprotein in a ovarian cancer: A molecular marker associated with multidrug resistance, *J. Clin. Oncol.* 3:311-315(1985).
63. Chan, H. S., Bradley, G., Thomer, P., et al, A sensitive method for immunocytochemical detection of P-glycoprotein in multidrug resistance human ovarian carcinoma cell lines. *Lab. Invest.* 59:870-875(1988).
64. Endicott, J. A., Ling, V., The biochemistry of P-glycoprotein-mediated multidrug resistance. *Annu. Rev. Biochem.* 58:137-171(1989).
65. Greenberger, L. M., Williams, S. S., and Horwitz, S. B., Biosynthesis of heterogeneous forms of multidrug resistance-associated glycoprotein. *J. Biol. Chem.* 262:13685-13689(1987).
66. Carlson, S. V., Till, J. E., and Ling, V., Modulation of drug permeability in Chinese hamster ovary cells., Possible role for phosphorylation of surface glycoprotein. *Biochim. Biophys. Acta* 467:238-250(1977).
67. Garman, D., Alberts, L., and Center, M. S, identification and characterization of a plasma membrane phosphoprotein which is present in Chinese hamster lung cells resistant to Adriamycin. *Biochem. Pharmacol.* 32:3633-3637(1983).
68. Hamada, H., Hagiwara, K., Nakajima, T., et al, Phosphorylation of the Mr 170000 to 180000 glycoprotein specific to multidrug resistance tumor cells: Effects of verapamil, trifluoperazine, and phorbol esters. *Cancer. Res.* 47,2860-2865(1987).
69. Chin, K. V., Tanaka, S., Darlington, G., et al, Heat shock and arsenite increase expression of the multidrug resistance(MDR1) gene in human renal carcinoma cells, *J. Biol. Chem.* 265:221-226(1990).
70. Chin, K. V., Udea, K., Pastan, I., et al, Modulation of activity of the promoter of the human MDR1 gene by Ras and P53. *Science* 255:459-462(1992).
71. Cornell, M. M., and Smith, D. E., A signal transduction pathway for activation of the *mdr1* promoter involves the proto-oncogene c-raf kinase. *J. Biol. Chem.* 268:15347-50(1993).
72. Mickly, L. A., bates, S. E., Richert, N. D., et al, Modulation of the expression of a multidrug resistance gene(*mdr1*/P-glycoprotein) by differentiating agents, *J. Biol. Chem.* 264:18031-18040(1989).
73. Burt, R. K., and Thorgeirsson, S. S., Coinduction of MDR1 multidrug-resistance and

- cycochrome P-450 genes in rat liver by xenobiotics. *JNCI.J.Natl.Cancer.Inst.* 80:1383-1386(1988).
74. Uchiumi,T., Kohno,K., Tanimura,H.,etal, Enhanced expression of the human multidrug resistance 1 gene in response to UV light irradiation. *Cell Growth Differ.*4:147-157(1993).
 75. Thorgeirsson,S.S., Huber,B.E., Sorrell.S., etal, Expression of the multidrug resistance gene in hepatocarcinogenesis and regenerating rat liver. *Science* 236:1120-1122(1987).
 76. Beck,W.T., Cirtain,M.C., Danks,M.K., etal, Pharmacological, molecular, and cytogenetic analysis of "atypical" multidrug -resistant human leukemic cells. *Cancer Res.* 47:5455-5460(1987)
 77. Danks,M.K., Schmidt,C.A., Cirtain,M.C., etal, Altered catalytic activity of and DNA cleavage by DNA topoisomerase II from human leukemic cells selected for resistance to VM-26 . *Biochemistry*,27: 8861-8869(1988).
 78. Danks,M.K., Yalowich,J.C., and Beck,W.T., Atypical multiple drug resistance in a human leukemic cell line selected for resistance to teniposide (VM-26). *Cancer Res.* 47:1297-1301 (1987).
 79. Cole,S.P., Bhardwaj,G., Gerlach,J.H., etal, Overexpression of a transporter gene in a multidrug resistance human lung cancer cell line. *Science* 258,1650-1654(1992).
 80. Krishnamachary,N., and Center,M.S., The MRP gene associated with a non-P-glycoprotein multidrug resistance encodes a 190-Kda membrane bound glycoprotein. *Cancer Res.* 53:3658-3661(1993).
 81. Morrow,C.S., Cowan, K.H. (1990): Gluthathione-S-transferase and drug resistance *Cancer. Cell* 2:15.
 82. Reed,J.C.(1995) Bcl-2:prevention of apoptosis as a mechanism of drug resistance . *Hematol Oncol Clin North Am*, 9:451.
 83. Dalton,W., Cress,A., Alberts,A., etal, Cytogenetic and phenotypic analysis of a human colon carconoma cell line resistant to mitoxantrone. *Cancer Res.*48:1882-1888(1988).
 84. Taylor,C.W, Dalton,W.S., Parrish,P.R., etal, Different mechanisms of decreased drug accumulation in doxorugicin and mitoxantrone resistant variants of the MCF-7 Human breast cancer cell line. *Br.J.Cancer.*63: 923-929(1991).
 85. Slovak,M.L., Ho,J.P., Bhardwaj,G., etal, Localization of a novel multidrug resistance-associated gene in the HT1080/DR4 and H69AR human tumor cell lines. *Cancer Res.* 53:3221-3225(1993).
 86. Zwelling,L.A., Slovak,M.L., Doroshov,J.H., HT1080/DR4:A P-glycoprotein-

negative human fibrosarcoma cell line exhibiting resistance to topoisomerase II-reaction drugs despite the presence of a drug-sensitive topoisomerase II.

JNCI.J.Cancer.Inst,82:1553-1561(1990).

87. Baas,F., Jongsma,A.P., Broxterman,H.J., etal, Non-P-glycoprotein mediated mechanism for multidrug resistance precedes P-glycoprotein expression during in vitro selection for doxorubicin resistance in a human cancer cell line. Cancer Res. 50:5392-5398(1990).

附录

附录 A

学位期间发表的文章:

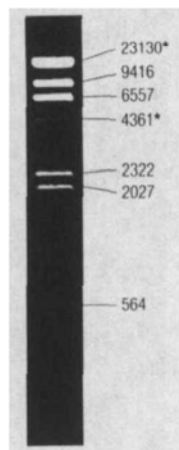
1. 刘勇 刘山鸣 姚俊杰等, 一个复合 STR 分离, 基础医学与临床 (已接受)。
2. 刘勇 刘山鸣 吴东颖等, 利用外切酶 III 建立次级亚克隆测序系统, 基础医学与临床 (已接受)。
3. 刘勇等, 综述: STR 与疾病 (待发表)。
4. 刘勇 刘山鸣 姚俊杰等, 一个复合 STR 分离, 第四届全国遗传学大会摘要。
5. 刘山鸣 刘勇 张小晓等, 河豚鱼 BAC 文库的构建, 中国医学科学院学报 1998, vol 3。
6. 刘山鸣 刘勇 张小晓等, Construction and Characterization of an Arrayed Pufferfish BAC library. Genomics (accepted)
7. 刘山鸣 刘勇 张小晓等, Construction and Characterization of an Arrayed Pufferfish BAC library. HUGO'98, March, Italy.

参与的工作:

1. 完成河豚鱼 α -TNF cosmid 候选克隆的杂交、挑选;
2. 完成河豚鱼与人 5p 一 exon 对应的 cosmid 候选克隆的杂交、挑选;
3. 协助完成河豚鱼与人 13 号染色体部分区域对应的 cosmid 候选克隆的杂交、挑选;
4. 协助完成河豚鱼 MDR BAC 候选克隆的杂交、挑选;

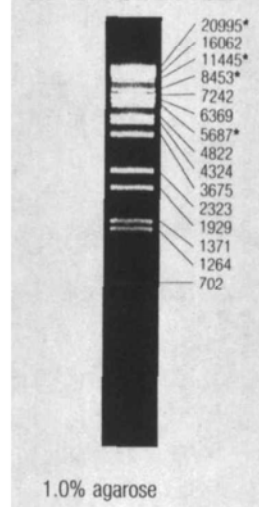
附录 B: 标准分子量

1. λ DNA/HindIII



1.0% agarose

2. λ DNA/BstEII



1.0% agarose

附录 C: 河豚鱼 cosmid 文库杂交膜说明

GENOMIC FUGU COSMID LIBRARY

About the library

The Fugu library provided by the Resource Centre was constructed by Elgar and Nizetic (unpublished data). It has an average insert size of 40 Kb and consists of approximately 76,000 clones. The vector used is Lawrist 4. More information about the library may be obtained from the Genome News, 20 (1995).

Production of high-density gridded filters

The library has been gridded in a 4x4 array on *four* 22 x 22 cm Hybond N and *one* 7.8 x 11.9 cm Hybond N⁺ nylon membranes (Amersham) using either a PBA Technology Flexys robot or a Genetix Qbot robot. The combination of Hybond N and N⁺ nylon membranes will not affect subsequent hybridisation if using radioactivity. Each clone has been spotted twice to give 36,864 clones (18,432 x 2) on the large membranes and 2,304 clones (1,152 x 2) on the small membrane. The filters have been labelled as follows:

Label	Plates gridded
FUGU1	Plates 1-48
FUGU2	Plates 49-96*
FUGU3	Plates 101-148
FUGU4	Plates 149-196
FUGU5	Plates 197-199

** plates 97-100 do not exist*

Clones spotted on the filters were grown overnight at 37°C on LB agar plates containing Kanamycin (30 µg/ml). The filters were subsequently processed by (a) SDS treatment, (b) denaturation and (c) neutralisation, then dried and crosslinked using a Spectronics Corporation XL-1500 UV Crosslinker. The filters are ready for pre-hybridisation in appropriate buffers.

User Information:

For successful removal of probes, and to enable re-probing, membranes **must NEVER** be allowed to **dry during, or after, hybridisation and washing**. Membranes can be stripped to facilitate re-probing by washing sequentially at RT in 0.2M NaOH for 30 min. and 0.1 x SSC / 0.1% SDS / 0.2M Tris-HCl pH 7.5 for 15 min. A gentler method, which extends the number of probings per filter, is to wash the probed membrane after use in a large volume of pre-hybridisation buffer overnight, prior to using another probe.

Subsequent publication:

Please acknowledge the originator of the library, Elgar and Nizetic, and the UK HGMP Resource Centre for providing the library in any publication arising out of using this resource. An example of how a clone (E.g. 171-m19) should be described in a publication is: 171m19 from the Elgar Fugu library.

INTERPRETATION OF HYBRIDISATION RESULTS

Each filter is labelled at the top left hand corner (FUGU1, FUGU2 FUGU5). This label also serves as an orientation mark. The four enclosed sheets (i, ii, iii and iv) provide detailed information about the layout of clones on each of 5 filters.

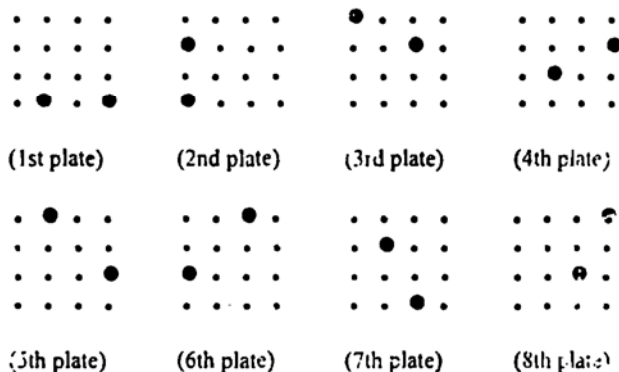
Sheet (i) shows a typical 22 x 22 cm membrane with 36,864 spotted clones. Each large filter is divided into 6 panels (indicated by ①, ②.....⑥). The panels are numbered from left to right, and eight 384-well plates are gridded, in duplicate, in each panel. Within each panel, numbers 1-24 correspond to the 24 columns and letters A-P are the 16 rows of the eight 384-well plates in the panel. Each dot represents a clone from a single microtitre plate well. A group of 16 dots corresponds to a 4x4 array of duplicate clones from the 8 microtitre plates (8 x 2). Thus each panel consists of 6144 spots (384 x 8 x 2). The figure at the top of sheet (ii) shows the order of the 8 plates (in duplicate) within each panel of filters Fugu1-Fugu4. Panel ① contains plates 1-8, ② contains 9-16 and so on. This plate order has been maintained on all the membranes, e.g.

Membrane FUGU1, Panel ①	Membrane FUGU1, Panel ②
plates 3 5 6 8	plates 11 13 14 16
2 7 3 4	10 15 11 12
6 4 8 5	14 12 13 13
2 1 7 1	10 9 15 9etc.

Sheet (ii) shows the filter (FUGU5). Since only 3 plates are left for this last filter a small membrane (7.8 x 11.9 cm) has been used to grid the remaining 3 plates.

The interpretation of positive signal(s) can be done by following the procedure given below:

1. Identify the membrane on which the positive signal is located (in duplicate).
2. Use the label position to orientate the filter and identify the panel in which the positive clones are located.
3. Work out the microtitre co-ordinates of the 4 x 4 array by referring to the row (A-P) and column (1-24) designations.
4. Within the identified 4 x 4 array determine which plate number contains the positive clone. Check that any two signals fall in one of the following patterns:



5. Record the identity of putative positives by following the convention PLATE-ROW and COLUMN (e.g. 154-L15)

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P																																																				
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100

(ii)

FUGU 1	
① 1 - 8 3 5 6 8 2 7 3 4 6 4 8 5 2 1 7 1	② 9 - 16 11 13 14 16 10 15 11 12 14 12 16 13 10 9 15 9
③ 17 - 24 19 21 22 24 18 23 19 20 22 20 24 21 18 17 23 17	④ 25 - 32 27 29 30 32 26 31 27 28 30 28 32 29 26 25 31 25
⑤ 33 - 40 35 37 38 40 34 39 35 36 38 36 40 37 34 33 39 33	⑥ 41 - 48 43 45 46 48 42 47 43 44 45 44 48 45 42 41 47 41

FUGU 2	
① 49 - 56 51 53 54 56 50 55 51 52 54 52 56 53 53 49 55 49	② 57 - 64 54 61 62 64 58 63 59 60 62 60 64 61 58 57 63 57
③ 65 - 72 67 69 70 72 66 71 67 68 70 68 72 69 66 65 71 65	④ 73 - 80 75 77 78 80 74 79 75 76 78 76 80 77 74 73 79 73
⑤ 81 - 88 83 85 86 88 82 87 83 84 86 84 88 85 82 81 87 81	⑥ 89 - 96 91 93 94 96 90 95 91 92 94 92 96 93 90 89 95 89

FUGU 3	
① 101 - 108 103 105 106 108 102 107 105 104 106 104 108 102 102 101 107 101	② 109 - 116 111 113 114 116 110 115 111 112 114 112 116 113 110 109 115 109
③ 117 - 124 119 121 122 124 118 123 119 120 122 120 124 121 118 117 123 117	④ 125 - 132 127 129 130 132 126 131 127 128 130 128 132 129 126 125 131 125
⑤ 133 - 140 135 137 138 140 134 139 135 136 138 136 140 137 134 133 139 133	⑥ 141 - 148 143 145 146 148 142 147 143 144 146 144 148 145 142 141 147 141

FUGU 4	
① 149 - 156 151 153 154 156 150 155 151 152 154 152 156 153 150 149 155 149	② 157 - 164 159 161 162 164 158 163 159 160 162 160 164 161 158 157 163 157
③ 165 - 172 167 169 170 172 166 171 167 168 170 168 172 169 166 165 171 165	④ 173 - 180 175 177 178 180 174 179 175 176 178 176 180 177 174 173 179 173
⑤ 181 - 188 183 185 186 188 182 187 183 184 186 184 188 185 182 181 187 181	⑥ 189 - 196 191 193 194 196 190 195 191 192 194 192 196 193 190 189 195 189

(iv) An alternative translation method is to use the table below. Identify the plate position of your positive clone, and then look up the plate number corresponding to that under the correct filter number and panel number. Well Id is determined as previously.

Filter No.	Panel No.	Positive Plate Position								
		1 st	2 nd	3 rd	4 th	5 th	6 th	7 th	8 th	
1	①	1	2	3	4	5	6	7	8	
	②	9	10	11	12	13	14	15	16	
	③	17	18	19	20	21	22	23	24	
	④	25	26	27	28	29	30	31	32	
	⑤	33	34	35	36	37	38	39	40	
	⑥	41	42	43	44	45	46	47	48	
	2	①	49	50	51	52	53	54	55	56
		②	57	58	59	60	61	62	63	64
③		65	66	67	68	69	70	71	72	
④		73	74	75	76	77	78	79	80	
⑤		81	82	83	84	85	86	87	88	
⑥		89	90	91	92	93	94	95	96	
3		①	101	102	103	104	105	106	107	108
		②	109	110	111	112	113	114	115	116
	③	117	118	119	120	121	122	123	124	
	④	125	126	127	128	129	130	131	132	
	⑤	133	134	135	136	137	138	139	140	
	⑥	141	142	143	144	145	146	147	148	
	4	①	149	150	151	152	153	154	155	156
		②	157	158	159	160	161	162	163	164
③		165	166	167	168	169	170	171	172	
④		173	174	175	176	177	178	179	180	
⑤		181	182	183	184	185	186	187	188	
⑥		189	190	191	192	193	194	194	196	
5			197	198	199					

cFugu Clones

The following clones in the cosmid Fugu library are known to generate false positive signals with many probes. They are either non-recombinant or 'noisy' clones. These clones have been kept in the library as they can be very useful as positive controls in the hybridisation process.

However, any clones requested from this list will not be sent.

5-F4	31-B12	84-A20	106-M4	127-F11	148-I20	170-C1
7-F1	33-H5	86-A6	117-B5	127-O16	152-C20	171-G16
12-E8	37-K17	101-E23	117-H12	133-D21	157-N22	173-P14
18-G16	38-A18	107-H21	117-D12	134-A3	158-G6	177-H9
22-J16	39-I19	107-K8	118-E6	134-A4	160-A5	180-F23
26-H11	63-G5	110-B21	118-N4	136-B18	160-C3	185-P10
27-E7	65-K20	112-B18	119-A2	137-E20	160-P10	191-G3
28-M17	76-F6	112-H24	119-B24	138-F20	164-B4	191-G4
30-L15	82-I5	112-O9	124-E11	139-I18	164-P3	198-B9

致 谢

本文是在恩师杨焕明教授、方福德教授、杨纯正教授的悉心指导下完成的。感谢恩师们给予我实验精心的设计、指导，从恩师们身上，我不仅学到了专业理论知识，更学到了为人的道理，这财富将伴随我一生，使我终生受益。

感谢郑超强教务长十年来对我的爱护、鼓励和无私帮助。

感谢美国华盛顿大学人类基因组计划测序中心的于军博士在测序过程中给予的无私帮助，您科学至上的学术精神使我懂得了科学真正的含义。

感谢罗会元先生对工作的关心、鼓励和支持。

感谢刘国仰教授在实验过程中给予的大力支持和鼓励。

感谢祝庆林老师在系列分析时给予的大力协助，您的助人精神令我感动。

这里，特别感谢我室的学兄、学友们，我们在实验中，在生活上的相互鼓励、相互帮助、相互劝勉，使我们一起走过了这一段不平常的历程，使我懂得友谊是成长的支柱，我将永远记住您们：吴东颖博士、刘山鸣博士、蔡胜和博士以及周巍同学、郎铁明同学、丁立同学。

感谢遗传室黄尚志教授、蔡淑媛老师、高春生老师、刘春云老师、马绍武老师、李淑华老师、郭玉凤老师、曾萱同学、戴理同学以及其他老师和同学对我工作、生活的帮助。

感谢教务处的诸位老师们长期以来对我的教育和帮助。

深深的感情，汇成一句：不负期望，学有所成，报效协和，终生无悔。

特别感谢我妻子、父母及家人对我学业的支持。