

# 基因起源中的精卫之谜 (下)

□ 撰文·供图/龙漫远

在加州大学查尔斯·兰格利的实验室，我的博士研究课题原本是验证中性进化论，可是“无心插柳柳成荫”，一个年轻的“精卫”基因来到我的面前，并将我带入了一个全新的研究领域……

## (八)

“带着精卫基因去发展您的科学生涯吧！这不是我的发现，而是您的创造。”

毕业在即，我的老师兰格利这样对我说。他说他没有预见到我的研究会朝一个新的学科发展，他没能指导我对这类问题的分析。我知道老师的

谦逊满含着对学生深切的期望。

不久，我收到8封来自美国一流大学的博士后邀请。其中，来自哈佛大学细胞生物系主任、诺贝尔奖得主沃尔特·吉尔伯特的信让我度过了一个不眠之夜。信是这样写的：

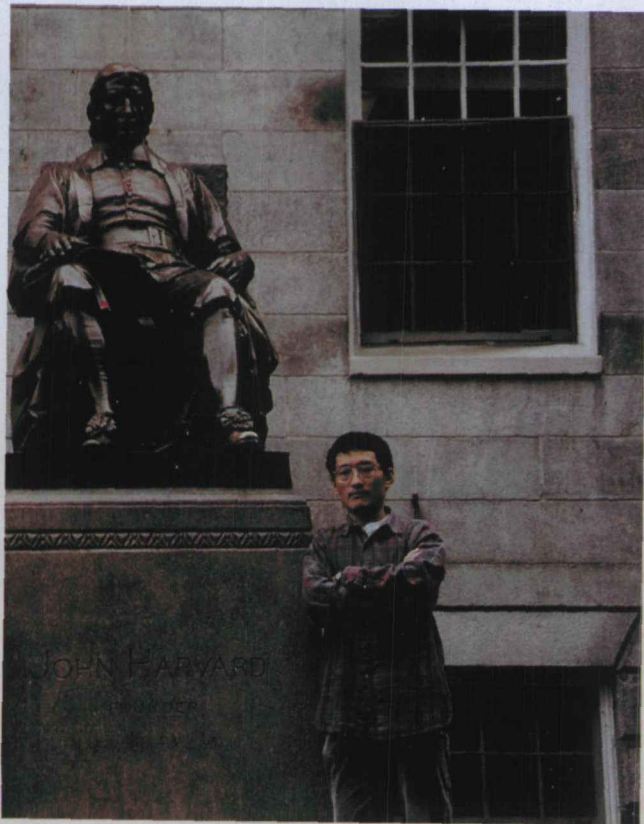
“我很高兴邀请您来我的实验室做博士后研究员。您可以继续做‘精卫’新基因的研究和其他与新基因结构有关的研究。”

吉尔伯特的经历在众多的西方科学大师中有着独特的传奇色彩。这是一位被《纽约时报》称为“天才”，被吉姆·沃森称为“一生中看到的最

关键的选择就那么几次，为了自己的理想，为了“精卫”，龙漫远毅然来到了哈佛大学。图为龙漫远在大学中央的约翰·哈佛铜像前留影。

智慧”的人物。他高中时的校长曾预言，吉尔伯特将会是一个“给我一根杠杆，我将移动地球”式的创造性人物。他1958年在剑桥大学拿到数学博士后，很快成为哈佛大学的理论物理教授。4年以后，正当他的理论物理生涯蒸蒸日上之时，他突然辞掉教职，投身于20世纪60年代初由吉姆·沃森等人发起的分子生物学的科学革命中。此后的十多年，他和沃森等人一同奠定了当代分子生物学基础，并于1983年获得了诺贝尔化学奖。20世纪80年代初，他开始研究遗传系统的进化，再一次离开了以前的领域——生物化学和传统分子生物学，并开始指导和创立了美国最大的生物工程产业，克隆了第一个人类乳腺癌基因。许多不同领域的研究人员都把他当成本领域的领袖，而不知道他同时还是其他领域的先驱。对我，吉尔伯特是一个类似荷马史诗主人公奥德修那样的人物。

但是，我有勇气成为这样一位科学大师的博士后研究员吗？他那出奇的怪想法和对身边研究人员的严格，与他在学术上的荣誉一样名满欧美生





物科学界。

“您想去他那里吗？那么在他面前，您必须所有的时候都优秀！否则您可能被一次错误所淘汰！”系里一位教授警告我。

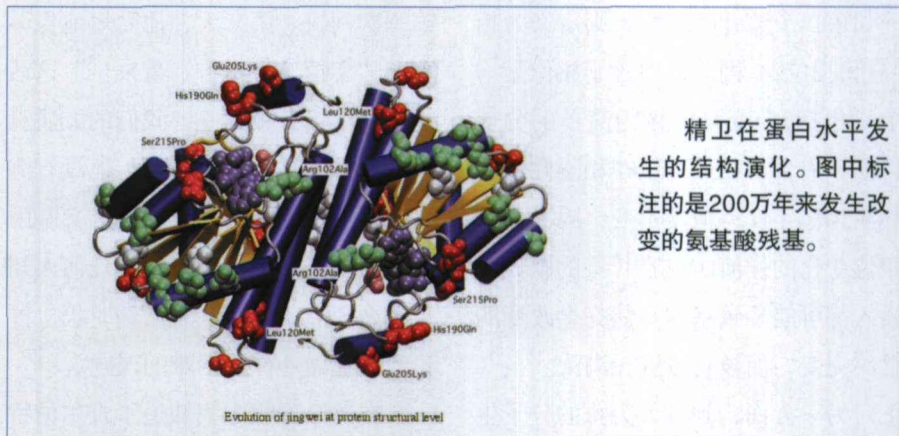
又是一次选择！而且可能是一次对事业关键性的选择。按我周围美国同事的看法，就我的情况而言，去哈佛是冒险。因为他们认为，我在博士生期间已经做出了第一流的工作，无论我去其他7个实验室中的哪一个，都会顺利做完博士后研究并找到教授职位。

这时，我读到著名华裔物理学家丁肇中的传记，他的经历激励着年轻人在学术生涯的发展中敢于冒险。我想，过去的蹉跎岁月已经浪费了我不少时光，如果再不冒险，往后将不会再有多少冒险和由此可能带来的发展机会了。“生于忧患，死于安乐”，走向平庸的舒适不是我们应当追求的。

“朝闻道，夕死可矣”，去听听智者的声音，是多么有诱惑力的一件事啊！更何况，在那里可以继续“精卫”基因的进化研究。于是，带着一箱子“精卫”研究的资料，我登上了去哈佛大学所在地波士顿的飞机。

### (九)

“精卫”基因的起源清楚地证明了由反转录产生的DNA片段与其他基因编码区的重组过程。这一过程的编码区（外显子）的重组，在20年前真



核生物基因内的干涉顺序（内含子）刚发现后，即由吉尔伯特所预料。但是由反转录产生的DNA参与重组过程则是吉尔伯特当时所未曾料到的。在他的早期理论中，发生在基因组序列水平上的非同源重组被认为是主要机制。但是无论重组过程的分子细节怎样，由不同基因的外显子重组以构成新基因结构则是吉尔伯特重组理论的核心。这一理论远远不同于大野·乾的基因重复理论。后者在概念上认为基因的每个核苷酸都重要，而前者则认为基因的外显子为最起码的重组功能单位，因而能解释已存在蛋白亚单位的有限性。

那么，由“精卫”所代表的基因结构形成的分子机制是否在真核生物新基因产生过程中普遍存在？这种机制是否在生命存在的早期就已经开始？

这是我在哈佛大学4年中研究的主要问题。我运用吉尔伯特实验室的专业设备，同时应用分子生物学实验和分子信息学的计算手段研究基因的起源。通过对几百万份DNA

序列的生物信息分析，我证明了外显子重组机制的普遍性。这是一个多学科交叉的领域，涉及分子生物学重组机制的实验、分子进化论、统计概率论和计算科学。下面我将着重描述分子生物学原理及对大规模分子统计行为的影响。

由于新基因的DNA序列进化速度很快（如“精卫”所示，快50~100倍），因此，若起源的基因进化年代比较久远，基因的重组关系常常变得不可鉴别。于是，我们工作的重点是寻找进化上保守的基因分子特征。我们发现最保守有用的分子特征，是基因的干涉顺序（内含子）的相位的生物信息，分析的结果，改变了许多传统的关于基因结构的观点。

内含子相位，是由麻省理工学院的诺贝尔奖得主、内含子发现者菲立浦·夏普定义的，它指内含子在基因内相对于遗传密码的3个核苷酸的位置。如果一个内含子是处在2个完整的密码间，则这一内含子定义为相位0；如果内含子是位于密码子内的第



一和第二个核苷酸之后,则定义为相位1和相位2。因此,内含子相位是一个只包括3个变量的极为简单的分子特征。内含子相位可以在相隔几亿年甚至10亿年保持恒定不变,因为导致相位变化的任何DNA小片段的缺失与插入都可能导致基因功能完全改变的移码突变,而被自然选择淘汰。

另一方面,到今天为止的分子生物学研究表明,内含子通常是没有功能的。因此,无论内含子的相位是什么,都会因为不影响基因功能而免于自然选择被保留下来。这就构成了内含子相位的随机统计分布预期的生物学依据。

要测验内含子相位的随机分布的预期结果,似乎是一件再简单不过的事。但实际的研究过程却是一场前所未有的考验。这是由于真核细胞生物,特别是人类的基因组中,存在大量的相似的重复顺序。不把所有的DNA相似序列剔除掉,任何分析都将带来不可估量的偏差。因为相似的DNA序列所含的相似的内含子相位,是由共同祖先基因起源的,不反映基因组的一般性质。但是,重复序列的剔除工作非常困难。首先,要求数亿次序列高敏感的比较。一开始,吉尔伯特和我都低估了技术方面的难度。我们设计了一项当时领域里最灵敏的方法,去寻找淘汰重复序列。半年紧张的技术努力,写出了几百页的结果报告。我以忐忑不安的心情把报告送给了吉尔伯特,等待着他的

第一次判决。第二天,他简短地告诉我:“忘掉这份报告,重来!”这话包括两个信息:第一,我们的方法没有达到他的高要求;第二,他没有淘汰我。但怎样大规模自动化定义基因家系,当时我们领域里没有任何人知道。两年后,夏普在《纽约时报》评论我们在做一件最困难的课题。

在这个过程中我知道了吉尔伯特培养青年研究人员的两个特点。第一,在设计课题时只考虑所研究的问题是否有重要意义,而技术可能性则不在考虑之列。第二,对进行研究的青年人在技术上提出很高的要求,让他们在“山穷水尽”的困难中奋斗,并在这一过程中选择或淘汰他们。

此刻,面对着吉尔伯特,这样一位严峻的、当代西方科学界最辉煌的智者之一,我开始懂得来哈佛之前同事们劝告的具体含义。去斯坦福,去芝加哥,去杜克,或去哈佛另外一个实验室,我也许会容易得多。但是,我已经选择了冒险。我只能想像纵身跃入结满浮冰的小河的体验,也许是不多的人所能体验的喜悦。在云贵高原上,我曾接受过大山的挑战。当然,在感情和意志的世界里,我最终还得在理智上找到目前技术挑战的答案。半年过后,经过无数次的验证,我再一次将报告送给吉尔伯特,他的脸上终于露出了少有的笑容。

在这些研究中,我证明以前领域里所有的有关猜想都是错误的。我发现了内含子相位在真核生物核基因组

的普遍非随机分布。更让人惊讶的是基因内含子相位的对称分布,即同种内含子的相依现象。这样的对称分布,揭示了作为“精卫”基因创造机制的外显子重组的普遍性。我们发现,至少一半以上的真核生物基因,曾经历了类似“精卫”基因的起源过程。这些工作,描述了第一幅关于基因起源的详细图景。于是,“精卫”基因起源所代表的分子机制从特殊到普遍意义都开始得到了证明。

## (十)

顺利完成了哈佛大学的博士后研究工作,在一个枫叶如丹的季节,我跨进了芝加哥大学的校门。在一幢哥特式的欧洲中世纪风格大理石楼房里,我建立了一个以研究基因起源与演化为中心课题的实验室。

深红、丹朱、橘黄的常春藤叶子在实验室外的大理石墙上热情地欢迎着我,但是,石墙上那些雕刻着的石像却在冷漠威严地注视着我。我知道,这些厚重的建筑和雕塑记录了芝加哥大学不寻常的历史和光荣,尽管我并不相信这是藐视和挑战后来者的理由。一位同事来到我年轻的实验室,告诉我在这个空间里,100多年来先后生活过几位开创和改变了演化生物学和遗传学的先驱人物,例如舍沃·怀特和我在哈佛大学另一位博士后导师理查德·陆文顿。他更指着实验室对面一幢灰色大理石建筑的地



下室，说：“你知道，那家伙曾在那里做出了一个实验，不仅改变了物理学，同时也改变了人类历史在20世纪的进程。”我知道，他在说费米，一位20世纪40~50年代的芝加哥大学教授，在地下室里建成了世界上第一座受控原子核裂变链式反应堆，为人类也为芝加哥大学留下了“盗火者”的足迹。

在这样不轻松的氛围与期望中，我开始了独立的科学研究生涯。精卫基因的研究给了这里的同事与管理层以信心，把我请来从事基因起源的第一个系统的研究。但基因起源接下来的一个问题却让我感到如芒刺在背：什么是精卫基因的生物学功能？搞清一个全新基因的功能不是一件一蹴而就的事。我记起了几位前辈在理解一个基因的功能上毕生奋斗而最终一无所获的故事，一位数学家在他的生命之灯熄灭之前，告诫他的儿子别像他那样终身为一道方程的证明而穷尽生命。

有一位同事安慰我不必介意精卫的具体生物学功能。因为，对演化生物学家而言，知道精卫是一个有功能的基因就够了，我们的任务是演化遗传学方面的理解。我知道他的好意，也知道他不是没有道理。选取容易的课题可能产生一个表面上多产的研究纪录，这对一个还没拿到终身教职的助理教授来说是一个很实际的选择。而更糟糕的是，在精卫基因功能研究上，含有这一基因的几个物种分布在



芝加哥大学校园内的石墙上雕刻的石像冷漠而威严，百年来，它们就这样注视着校园内人们来来往往、世事变迁。



西非的热带雨林之中，有关的生物学信息所知甚少，果蝇分子生物学家们所创造的分子生物学工具基本派不上用场。寻求容易的课题是一个诱惑，一个智者的声音却在告诫我，“您不能在科学这块美丽的木板上钻许多小孔，那不是科学发展的途径和一个真正的科学家应该走的路”。毕竟，作为第一个新基因起源研究的系统，搞清精卫基因的功能对整个基因系统演化的理解至关重要。是的，对一个登山者而言，要他半途下山不是一种光荣。于是，我对科学前辈们的执著有了更新的一层理解——那是他们的精神生活，也是他们的责任。大山立在那里，作为登山者，您怎么可以无动于衷，甚至想当逃兵呢？如果课题失败而拿不到终身教职又有什么关系，云贵高原上那第一次“土插队”的涧水长流、绿竹遍野的地方，不是也可以活人吗？

有了这些科学探险者的决心，利用芝加哥大学的智力环境，我很快组

织起一支志同道合的研究团队，开始就精卫的功能及功能演化的过程进行探索。

#### (十一)

背水为战的智力和意志的角逐开始了。两年过后，系里一位好心的同事开始关注我的课题走向，他不希望我失败而离开芝加哥大学。但此时，没有任何一个课题有一个明确的有希望的结果，我们遇到了前所未有的困难。精卫基因功能的生物化学研究遇上了从未想像的现象：精卫能在大肠杆菌中大量表达，但就是不能纯化，也就无法进行下一步酶活性的测定。一个月，两个月，半年，一年……我们停在那里，我们用尽了所有的办法。我的合作伙伴，明尼苏达大学生化系的托尼·丁教授为我们担心，劝我们改课题。可我知道，这场战争还没有到最后一刻，“两军相逢勇者胜”，我和我的学生们顽固地坚持了



下去。

第三年初,转机开始出现。首先是博士后研究员王文发现了精卫母基因“炎帝”的完全外显子—内含子结构,同时他发现果蝇的近缘种中存在新基因的信号。其次是博士研究生张建明发现“炎帝”基因存在许多稀有密码,从而使大肠杆菌的翻译机制不能识别这些密码子。他推测,大肠杆菌一定在合成精卫的多肽链时放进了错误的氨基酸残基,从而改变了精卫蛋白分子本应有的三维结构,导致纯化过程无效。同时,博士后研究员埃瑟尔·贝特伦和博士研究生凯文·桑顿还发现了新基因在演化过程中在性染色体和常染色体间呈方向性移动。博士后研究员安娜·尤娜帕等人则发现精卫的内含子—外显子结构正在经历一个适应性演化过程,而博士后研究员弗雷德里克·布鲁勒在一个新发现的非洲果蝇物种中也找到了精卫基因的存在。

在美国西北大学药物及生化系的教授,我的妻子李丽明(她是生物化学博士,我戏称之为有执照的生物化学家)的帮助下,张建明找到了细菌基因组的修饰表达技术,解决了由“炎帝”基因的稀有密码带来的问题。这一突破,实现了精卫酶蛋白纯化制备的顺利完成,为进一步研究精卫的生物化学功能打下了基础。我们看到,精卫基因在200万年的过程中,演化出两个不同于母基因乙醇脱氢酶的新生物化学功能。在酶反应中

精卫选择长链醇类底物,并且演化出催化外激素和保幼激素的代谢控制功能。这些精卫酶生物化学功能测定试验结果,让我们不得不推测,精卫在演化过程中适应性地改变了蛋白质顺序从而改变了酶反应的底物。这一推测在我们的蛋白质三维结构演化分析中得到了证明。

运用瑞典卡罗林斯卡研究所测定的两个分离8000万年的远缘果蝇物种的乙醇脱氢酶三维结构,我们推测出原始精卫的乙醇脱氢酶结构域的三维结构。然后,比较精卫基因DNA序列的种间差异,我们鉴别出精卫在过去200万年间发生的蛋白质序列的演化,并把这些序列变化的位置和类型定位在原始精卫的三维结构上,从而推测精卫酶结构和功能的演化。我们看到,两个氨基酸演化替换刚好发生在酶底物分子进入酶反应中心的进口。卡罗林斯卡的科学家过去的研究已经证明,改变这一进口会导致酶对底物的选择。

至此,我们知道了,新的蛋白质生物化学功能最近还在自然界产生并不断地演化,今天我们为之赞叹不已的蛋白质家系及其功能的巨大多样性和精巧结构,如同生命在其他层次所表现的魅力迷人的多样性一样,都有一个不断演化和起源的过程。

## (十二)

精卫基因新的蛋白质功能的发现

及其母基因“炎帝”的鉴别,揭开了第一个新基因系统起源的完全过程。2004年,在精卫的新功能发表在美国国家科学院院刊几个月之后,演化生物学领域内最好的教科书之一《进化》完整系统地介绍了精卫的起源过程及其新功能。我佩服该书作者道格拉斯·富图玛的严谨和耐心,11年前精卫基因首次被发现,在其他教科书都已经开始介绍时他没有动心。11年后当我们用实验证明了“精卫”的功能,证明了精卫起源机制的普遍存在,他才将其收入教科书,传学后人。这是学生们和合作朋友们的努力和贡献。我为我们在这11年里没有放弃而高兴。

11年的努力,同前辈中许多创造性活动相比,是微不足道的。两年前去西班牙巴塞罗那讲学时,我的博士后研究员贝特伦带我参观了巴塞罗那市著名建筑家安东尼奥·高迪设计的一幢教堂建筑,圣家堂(Sagrada Familia)。埃瑟尔告诉我,圣家堂已经建设了120多年了,计划在100到200年后完成。今天的圣家堂已是宏伟新奇之极,我不能想象建筑全部完成后该是何等的景象。如果这只是巴塞罗那市民宗教的热诚,俄亥俄州立大学和其他大学的几代玉米遗传育种家们坚持100多年作好一个重要课题,最近完成的玉米百代含油量的选择试验则显现了一个长远的研究目标和为完成这个目标所作的百年不懈的努力。在波士顿一个朋友家里,我看



到记录这些珍贵的试验记录和遗传分析的手稿。这百年记录告诉了我们改变20世纪世界农业科学研究和生产历史的美国玉米遗传育种获得巨大成功的秘密，嘲笑了在科学研究中急功近利眼光短浅的人。

兰格利教授专门从加利福尼亚写了一封热情洋溢的祝贺信。信中说“我知道这一研究的难度，我为你们的成就高兴。请把我真诚的祝贺转告张建明”。紧接话锋一转，“希望你们最终能找到精卫基因的表现型”。那是又一座需要攀登的高山峻岭。我的视线投向办公桌上兰格利老师11年前送我的毕业贺卡，上面是一位登山运动员正在峭崖上攀援，脚下是万丈深渊，头上是直刺蓝天的壮丽峰顶。贺卡后面的赠言墨迹还没有褪色：“漫远，从今一别，我将在远处看着你。”

江之永矣……

### (十三)

那年夏天，我又回到云贵高原那个流水潺潺、绿叶葱茏的山村。我默默地向那位已经不能再见的老人讲述了“精卫”复生变成小鸟以后的故事。我告诉他，这故事比他当年讲的传说长多了，东方和西方的神话传说都成了这故事的一部分。“精卫”象征着一个新的科学探索领域的诞生。但是我们所看到的，还只是这个美丽而令人不可思议的世界之一隅。有待理解的世界在膨胀，众多的谜底还未

解开，新产生的理论与技术问题不断构成挑战。

第一个基因或基因组在生命起源中是怎样产生的？

新基因是怎样地把一个新的分子功能加进一个自然界长期演化历史形成的功能系统及其基因控制网络中并进而改变这一系统功能的？

最近产生的新基因都给物种带来什么样的功能和表型？

新基因在基因组中的起源会构成一个不断膨胀的基因组或是一个平衡的分子世界？

新基因的产生是一个中性的随机过程还是自然界环境变迁所导致的选择结果？

新物种的产生或物种的生存是以新基因的产生为条件的吗？

我们之成为人，是因为人类的祖先在走出非洲后的短暂进化中，产生

了独有的新基因，并由此产生智慧创造了灿烂文明？

新基因的起源知识在医学、农业甚或人类文明进程中的前景如何？

……

我们现在还不知道这些重要问题的答案。但是，精卫的发现表明这样的问题并不是不可以研究的。探索的结果将会对生命最基本单位之一的基因的起源、意义及应用前景产生深层次的理解，这将是一部由我们和未来的生命科学工作者谱写的新的篇章。■

#### 作者简介

龙漫远，博士，美国芝加哥大学终生教授和北京大学长江讲座教授。其发现的“精卫”基因第一次从分子水平揭示了基因的起源过程，并由此开创了一个新的学科方向。

(责编 邹星)



龙漫远教授和他实验室的研究生、博士后、访问学者们在芝加哥大学的实验室里。他们在这里为揭开基因起源的秘密不懈地努力着。