



点 评

DNA 测序技术引领中国基因组科学走向未来

吴佳妍[†], 肖景发[†], 张若思, 于军*

中国科学院北京基因组研究所, 中国科学院基因组科学及信息重点实验室, 北京 100029

[†] 同等贡献

* 联系人, E-mail: junyu@big.ac.cn

收稿日期: 2010-11-25; 接受日期: 2010-11-29

1 人类基因组计划 10 年后, 全球科学家对基因组研究的看法

10 年前, 人类基因组草图完成, 全世界都在期待着这一计划能最大限度地对人类疾病起到革命性的影响(2000 年 6 月 26 日, 美国总统比尔-克林顿在白宫的讲话). 然而, 10 年后, 人们一直期待的人类疾病的革命依然未能实现. 但无论如何, 人类基因组计划的完成以及目前测序技术的创新发展, 是生物学上的一场革命. 它改变了很多生物学家的职业生涯, 鼓励他们去解决新的生物学问题、面对新的挑战.

最近, *Nature* 做了一项非常有意义的调查, 超过 1000 名生命科学研究领域的科学家都参与了调查 (<http://go.nature.com/3Ayuwn>). 调查表明, 几乎所有生物学家都在一定程度上受到人类基因组计划的影响. 69% 的受访者表示, 是人类基因组计划刺激他们成为一名科学家或者改变了他们的研究方向. 接近 90% 的人认为, 他们自身的研究获益于人类基因组被测序, 其中 46% 的人认为这种影响是“巨大的”. 同时, 有接近 1/3 的人几乎每天都在他们的研究中使用到基因组. 一位受访者说: “对于像我这样的年轻研究者, 很难想象没有基因组, 生物学家将如何实施工作.”

面对自身基因组信息急剧增长所带来的各种变化和可能性, 受访者的回答则不尽相同. 当被问及如果你可以对任何感兴趣的内容进行测序时你会测些什么时, 很多受访者都列上自己的、孩子的或者其他

家庭成员的基因组(有些人的答案中还包括他们的宠物猫或狗). 约 13% 的人已经对自己的 DNA 进行了部分测序和分析. 当人类全基因组测序费用达到 1000 美元左右时, 1/5 的人将对自己的基因组进行完整测序; 大约 60% 的人则期待着费用能够在 100 美元左右或者免费; 而其他人考虑的则远比测序自身基因组更多, 他们明确表示: “即使有人付钱给我, 我也不同意这么做.”^[1]

在受访科学家中取得广泛共识的是, 阐释和解析基因组远比获得基因组要面临更大的挑战, 目前基因组应用的主要障碍是依然缺乏对于基因组生物学最基础的理解. 大多数人认为, 人类基因组计划的实施所引发的针对解决人类疾病的革命, 在未来几个 10 年中是很难实现的. 毋庸置疑, 过去几十年的研究已经揭示出基因组的复杂性, 并且几乎涉及到人类学各个方面, 已经远超过先前的设想^[2]. 与复杂性直接相关的, 举例来说, 目前确认的 20000 个左右的编码蛋白质的基因, 占基因组 1.5%. 但仅仅知道这些, 并不足以解释这些基因的功能. 在后基因组时代的 10 年中, 更多的注意力被转移到蛋白质编码基因以外的区域, 发现很多非编码基因似乎拥有关键性的功能, 能够通过调控蛋白质编码基因的表达水平和产生大量的非编码 RNA 分子. 科学家通过 DNA 元件百科(the ENCYClopedia of DNA elements, ENCODE)和人类表观遗传基因组计划等大型项目的实施, 努力理解基因组的化学修饰, 目前被认为是控制基因调控的主要手段.

2 第二代测序仪的发展及影响

在过去的 30 年里,作为最重要的生物医学研究手段之一, DNA 测序技术自发明以来就一直在推动分子生物学发展方面起着至关重要的作用^[3]. 从早期 Frederick Sanger 的手工测序,以及基于 Sanger 法开发的第一代自动化测序仪,到目前的下一代测序平台,该领域已经发生了巨大的变化^[4]. 有人甚至将基因组测序技术的发展与半导体技术的发展相提并论,这也不无道理^[5]. 在过去的数十年中,每经过几年,测序速度就呈指数增长,与半导体工业发展的摩尔定律(Moore's Law)非常相似^[6]. 这种高速的发展,从根本上改变了人们研究所有生命蓝图的方式,推动了基因组学及其分支乃至其他密切相关学科的创立与发展,诸如比较基因组学、生物信息学、系统生物学以及合成生物学. 某种程度上, DNA 测序技术的进展已经使生命研究的基本元素发生了转变——从单一、局部的基因或基因的片段转变成整个基因组. 相反,这种转变又需要更加强大的测序技术来支持. 测序技术与其应用之间的协同关系使得两者的发展在可预见的未来内保持这种趋势,并且由于其对个体化疾病诊断与治疗具有确定性的推动作用而加速^[7].

在此力量的推动下,基因组学在人类基因组计划精细图完成后极速发展,大量全基因组计划纷纷完成,尤其是大型植物和动物基因组的增加. 如图 1 所示,病毒全基因组在 2002 年仅完成了 84 例,到 2010 年已经完成 1163 例;细菌全基因组在 2002 年完成 16 例,到 2010 年已经完成 93 例. 而真核生物的增长速度更是惊人(图 2). 真菌全基因组在 2002 年仅完成 4 例,在连续多年每年 10 例以上的增长量的推动下,2010 年已经达到 141 例. 在大型基因组方面,植

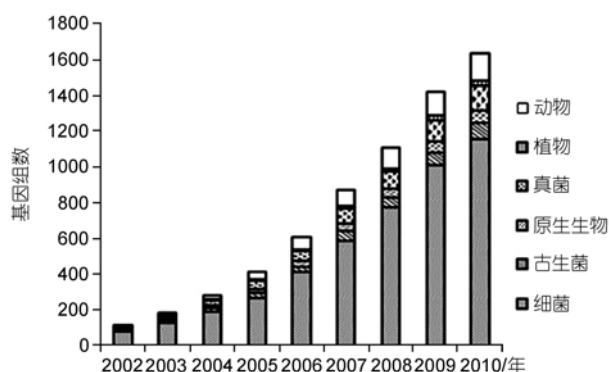


图 1 2002~2010 年每年测得的原核和真核生物基因组数目

物全基因组在 2010 年的完成量(32 例)是 2002 年(3 例)的 10 倍,其中增量最明显的是 2009 和 2010 年,分别是 11 和 8 例. 在更多的大型基因组测序完成过程中,下一代测序技术贡献的力量是不可回避的. 令人震惊的是,动物全基因组的完成速度,从 2002 年的 8 例到 2010 年的 151 例,整整增长了 19 倍. 可见,10 年前穷尽多个研究所几年之力,耗费大量资源才能完成的大型植物动物基因组,在 10 年间发生了天翻地覆的变化,其完成速度、完成质量、消耗资源都有了非常大的改善.

仅仅在过去 12 个月的时间里,就有苹果基因组、青鸟基因组、熊猫基因组、黄瓜基因组和玉米基因组等大型基因组得以完成^[8~12]. 这些全基因组的获得,虽然在使用第二代测序仪进行测序的同时,还常依赖第一代测序仪所产生的长片段序列,但测序效率的提高是有目共睹的. 非常有意义的是,由中国华大基因研究中心完成的熊猫基因组,在完全依赖短片段序列的测序结果的基础上,完成了熊猫基因组草图的构建.

DNA 测序技术 30 年的创新和发展开创了基因组测序的新时代,从手工的一次一个样品发展成为基于大规模阵列的高度自动化技术. 测序通量呈指数提高,而成本急剧降低. 1000 美元人类基因组的目标将变得更加现实. 快速、廉价的测序能力将引领人们开辟比较基因组学分析、疾病诊断以及个性化(个体化)医疗等新领域.

3 中国科学家在全球基因组学领域的地位和作用

正如陈竺和赵国屏先生在《中国科学: 生命科

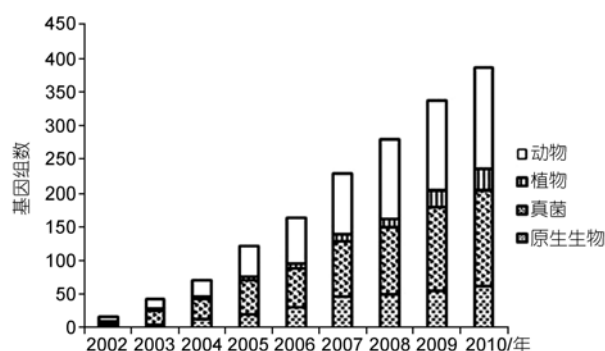


图 2 2002~2010 年每年测得的真核生物基因组数目

学》上发表的文章“基因组学在中国——从计划到实践的 10 年”中总结出中国基因组科学的 10 年来发展的历史^[13]。在人类基因组计划初期, 中国“国家高技术研究发展计划(863 计划)”专家组彻底评估了与人类健康和疾病相关的国际研究趋势, 最终建立了 2 个 1% 的目标, 分别指向基因组测序和全长 cDNA 识别。为了完成这项重大计划, 国家建立了上海和北京国家人类基因组中心, 这两个国家中心在人类基因组计划中起到非常重要的角色。该计划对基因组学在中国的发展发挥了不可估量的作用。此外, 计划的设计还包括加强疾病基因组的研究, 重点放在通过收集中外人类遗传资源产生的疾病相关基因的克隆和识别的相关网络来建立疾病模型。南方基因中心主持的这一疾病基因组计划在 1998 和 2004 年获得“国家重点基础研究发展计划(973 计划)”的连续支持。在中国, 以人类健康为导向的功能基因组研究的相关领域包括生物信息学、转录组学、蛋白质组学、结构基因组学和其他技术平台等, 都与人类基因组计划一起迅速发展起来。

两个国家重大科学研究计划——人和水稻的基因组学研究以及一系列基因组测序和功能基因组分析一起, 对中国的生命科学研究和生物技术的发展做出了前所未有的贡献(表 1)。特别是人和水稻基因组研究为后来的基因组学在中国的发展奠定了基础, 同时树立了中国基因组科学在国际上的重要地位。在此基础上, 中国科学院北京基因组研究所、南方基因中心和华大深圳研究院等构成了国内基因组数据资源的主要诞生地。利用人类基因组研究技术和资源优势, 北京基因组研究所和华大基因研究中心等机构在人类基因组计划后期开展了“超级杂交水稻基因组计划”和昆虫(家蚕)基因组测序的相关研究。“超

级杂交水稻基因组计划”是由国家杂交水稻研究中心、华大基因研究中心和中国科学院遗传与发育生物学研究所植物生物技术开放实验室三家单位联合开展的研究项目, 旨在通过获取中国高产超级杂交水稻的基本遗传信息, 开发包括丰产、优质、抗病、耐逆、育性、成熟期有关的资源和功能等重要农艺性状的基因。中国家蚕基因组计划是中国科学家在完成人类基因组 1% 测序工作、水稻基因组“框架图”和“精密图”之后, 向人类贡献的第三大基因组研究成果。这两大计划的实施和完成也标明了中国基因组学的自主研发能力。南方基因中心在 2009 年完成了血吸虫基因组计划, 并将日本血吸虫的基础序列数据通过上海研发公共服务平台科学数据共享网站, 向全世界公布^[13], 表明中国基因组信息技术在与国际接轨方面, 迈出了重要的一步。2008 年, KACST(King Abdulaziz City for Science and Technology)与中国科学院北京基因组研究所合作开展另一项重大植物基因组研究计划——“椰枣基因组计划”, 该计划是中国生命科学领域科学家参与的由国外政府资助的最大规模的国际科学研究计划, 该计划的签署和实施标志着中国在国际基因组学中的重要地位。近日, 中国科学院北京基因组研究所/中国科学院上海生命科学研究院植物生理生态研究所通过与中国水稻研究所、中国科学院遗传与发育生物学研究所、美国密歇根州立大学、华中农业大学和美国康奈尔大学等单位合作, 结合第二代测序技术和自主开发的基因型分析方法构建了高密度的水稻单体型图谱(HapMap), 并对籼稻品种的 14 个重要农艺性状进行全基因组关联分析, 确定了这些农艺性状相关的候选基因位点。相关研究成果于 2010 年 10 月 24 日以全文形式发表在 *Nature Genetics* 上^[21]。另外, 深圳华大研究院以企

表 1 中国机构主要承担和参与已完成的动植物基因组测序项目

物种	国内主要承担机构	文献
Part of <i>Homo sapiens</i> (人)	中国科学院华大基因研究中心等	[14]
<i>Oryza sativa</i> (indica cultivar-group) indica(水稻)	中国科学院华大基因研究中心等	[15]
<i>Bombyx mori</i> (家蚕)	西南农业大学、中国科学院北京基因组研究所和华大基因研究中心等	[16]
<i>Gallus domesticus</i> (家鸡)	中国科学院北京基因组研究所和华大基因研究中心等	[17]
<i>Homo sapiens</i> (人)	深圳华大研究院等	[18]
<i>Schistosoma japonicum</i> Anhui isolate(血吸虫)	南方基因中心等	[19]
<i>Cucumis sativus</i> L(黄瓜)	深圳华大研究院等	[11]
<i>Ailuropoda melanoleuca</i> (大熊猫)	深圳华大研究院等	[10]
<i>Harpegnathos saltator</i> (蚂蚁)	深圳华大研究院等	[20]

业模式成功开展了一系列基因组研究计划,如国际大熊猫基因组计划、黄瓜基因组计划和蚂蚁基因组计划等,特别是国际大熊猫基因组计划开创了第一次采用第二代短序列测序方法完成大型动物基因组测序的先例,该研究成果成为 *Nature* 当期的封面文章^[10]。除上述成果外,中国还有许多大型动植物基因组计划在实施和酝酿中,如鲤鱼基因组计划、千人基因组计划等。总之,虽然中国在基因组测序领域起步较晚,但已经跟上了世界功能基因组研究的步伐,所取得的成绩显示了中国生命科学研究和生物技术的发展。

国际人类基因组计划在完成了测序和 HapMap 计划之后,开始进入对人类疾病和人类表型进行相关研究的攻坚阶段。中国的基因组学研究也处于新的历史

阶段,伴随着新的高通量生物技术的不断涌现,生物信息数据急剧增长,形式日趋多样,如何更好地对海量数据进行管理和深度挖掘是解读这些信息的关键。基因组学数据的积累,同时推动了生物信息学的发展,这两个“孪生”学科成为推动生命科学发展的重要动力。功能基因组学、蛋白质组学、结构基因组学乃至代谢组学、基因组药物学等领域的发展,越来越依赖于整合的生物信息数据库、基因组信息分析挖掘方法和生物计算软件分析工具等。生物信息资源是关系到国民经济和社会可持续发展的重要战略资源,以数据挖掘为核心的生物信息技术则是生物信息资源能否被充分利用、生命科学研究能否取得突破的技术关键。在承接良好的开端和基础上,期待中国基因组学能够在 DNA 测序技术飞速发展的引领下走向更加辉煌的未来。

参考文献

- 1 Declan Butler. Science after the sequence. *Nature*, 2010, 465: 1000—1001
- 2 Hayden E C. Human genome at ten: life is complicated. *Nature*, 2010, 464: 664—667
- 3 Gilbert W. DNA sequencing and gene structure. In: Forsén S, ed. Nobel Lectures in Chemistry 1971~1980. 1980. Singapore: World Scientific Publishing Co, 1993. 408—426
- 4 Sanger F, Coulson A R. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol*, 1975, 94: 441—448
- 5 Shendure J, Mitra R, Varma C, et al. Advanced sequencing technologies: methods and goals. *Nat Rev Genet*, 2004, 5: 335—344
- 6 Moore G E. Cramming more components onto integrated circuits. *Electronics*, 1965, 38: 114—117
- 7 周晓光, 任鲁凤, 李运涛, 等. 下一代测序技术: 技术回顾与展望. *中国科学: 生命科学*, 2010, 40: 23—27
- 8 Velasco R, Zharkikh A, Affourtit J, et al. The genome of the domesticated apple (*Malus × domestica* Borkh.). *Nat Genet*, 2010, 42: 833—839
- 9 Warren W C, Clayton D F, Ellegren H, et al. The genome of a songbird. *Nature*, 2010, 464: 757—762
- 10 Li R, Zhu H, Ruan J, et al. The sequence and de novo assembly of the giant panda genome. *Nature*, 2010, 463: 311—317
- 11 Huang S, Li R, Zhang Z, et al. The genome of the cucumber, *Cucumis sativus* L. *Nat Genet*, 2009, 41: 1275—1281
- 12 Schnable P S, Ware D, Fulton R S, et al. The B73 maize genome: complexity, diversity, and dynamics. *Science*, 2009, 326: 1112—1115
- 13 陈竺, 赵国屏. 基因组学在中国——从计划到实践的 10 年. *中国科学: 生命科学*, 2008, 38: 879—881
- 14 Lander E S, Linton L M, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature*, 2001, 409: 860—921
- 15 Yu J, Hu S, Wang J, et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). *Science*, 2002, 296: 79—92
- 16 Xia Q, Zhou Z, Lu C, et al. A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*). *Science*, 2004, 306: 1937—1940
- 17 Hillier L W, Miller W, Birney E, et al. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*, 2004, 432: 695—716
- 18 Wang J, Wang W, Li R, et al. The diploid genome sequence of an Asian individual. *Nature*, 2008, 456: 60—65
- 19 Zhou Y, Zheng H, Chen Y, et al. The *Schistosoma japonicum* genome reveals features of host-parasite interplay. *Nature*, 2009, 460: 345—351
- 20 Bonasio R, Zhang G, Ye C, et al. Genomic comparison of the ants *Camponotus floridanus* and *Harpegnathos saltator*. *Science*, 2010, 329: 1068—1071
- 21 Huang X, Wei X, Sang T, et al. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat Genet*, 2010, 42: 961—967