

我们都有一个家

数万年前,现代亚洲人的一支,身上带着O3型的染色体,开始向北迁徙;数千年前,O3-122这一支种小米起家的人,开始向东向南扩展,最终遍布中国大地。■主笔 黄艾禾



9月初,是新学期开学的日子。复旦大学校园里,到处是报到的新生,复旦大学生命学院现代人类学研究中心的谭婧泽老师则忙着带几个助手给新生做体检。这不是普通意义上的健康检查,她是要配合课题,对学生们进行体质人类学的测量,取得一批数据。“学生们来自全国各地,正好是不错的数据来源”,而建立庞大的全国范围的人群数据库,是复旦现代人类学研究中心近年来一直孜孜以求的工作。

复旦的新生们以来自江浙一带的为多,但也有许多学生从云南、江西、河南、山东等省考来。而现代人类学研究中心的研究团队,主任金力是上海人,李士林是吉林人,其他同事分别来自新疆、甘肃、江苏、湖北……谭婧泽自己是宁夏人。

谭婧泽是18岁就离开宁夏的。到上海之前,她先到北京,在那里上大学和读研究生。毕业后,她又到了上海自然博物馆,再到复旦。从宁夏到上海,相隔“八千里路云和月”的故乡,已成为户口本或登记表上的“籍贯”,而对于她的下一代,宁夏只是母亲的老家,是



图/华盖

自己的上一代向上海迁徙的一个源头。

那么再上一代呢?谭婧泽的父母都是宁夏人。父母的父母呢?他们怎么到宁夏的?再一代代追上去,他

们的祖先又是哪里来的呢?——这个问题,总有孩子向自己的父母问起,一代一代问了好多年。

我们在学校读过的历史教科书,告诉我们人是从猿进化而来,进化是从数百万年前开始的。我们也都知道,北京的周口店曾出土了生活在60万年前的“北京猿人”,似乎顺理成章的,我们都是“北京猿人”的后代。

如果问今天的中国人是从哪里来的,我们也都知道黄帝大战蚩尤的传说,知道中华民族有五千年的文明史。但是,具体来说,上古的猿人是怎么变成今天的中国人的?我自己又怎么成为今天的我的,会呆在这个地方?这中间的环节又是怎样一环环连接起来的?

从前的人们,有几种途径来回顾自己的来路:或是从洪荒年代口口相传下来神话传说;或是古墓遗址中挖掘出的文物器物及化石;或是前人留存下来的古籍记录。古籍记录非常有限,神话传说又太玄虚,而文物出土常常支离破碎,使我们对于我们的前世总是感觉扑朔迷离。

最近20多年,人们又多了一条认识自己的路径:分子遗传学,从DNA上找自己的源头。

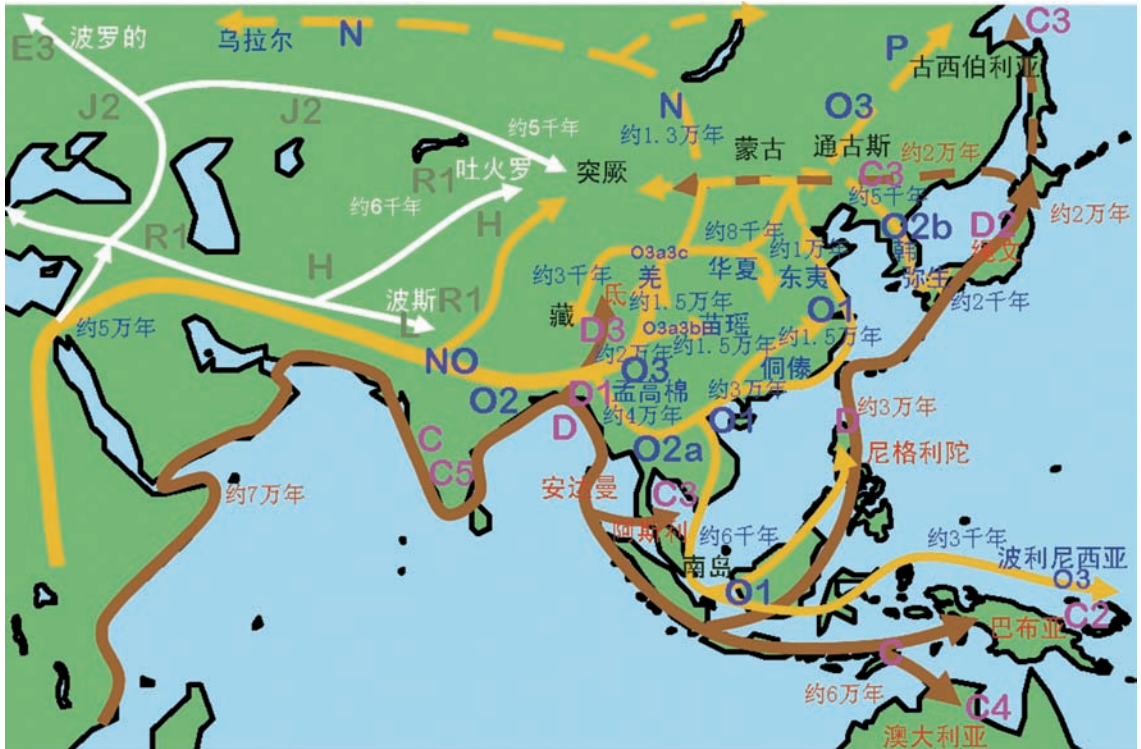
画一张亚洲人起源迁徙的地图

1998年,中国医学科学院医学生物学研究所褚嘉祐教授等人的一篇文章《中国人群的遗传关系》,在《美国科学院学报》上发表,在中国科学家中首先提出现代亚洲人类可能起源于非洲。

现代亚洲人类,当然包括中国人在内。也就是说,亚洲地区不存在从一两百万年前的直立人到现代人类的连续进化过程,“北京人”不是中国人的直系祖先。文章中的观点在今天看来都具有爆炸性。现代人类起源于非洲,这在世界的其他地方近年来已经大致被学术界接受(当然还不时有一些不同的质疑之声),但在中国,尤其是在古人类学领域,一种观点仍然持“多地区起源说”,也就是说,中国人是从东亚地区已经发现的那些古人类化石的后代,是独立起源的。

“一想到今天,这篇文章已经发表10年了”,褚嘉祐回忆当年的一幕,“那篇文章,本来是想投《科学》的,结果《科学》觉得这个结论太大,想放一下,谈家桢院士就建议投给《美国科学院学报》。发表后,《自然》做了长篇转载。还发了一个评论。”

亚洲现代人类迁徙地图



▲“国际Y染色体命名委员会”把全世界的Y染色体分为从A到R的若干大的类型，而中国人所带的，主要是O型、D型，其中D型比较古老，是棕色人种留下的基因，而O型分为O1、O2和O3三种，O1型从越南、广西方向进入中国，沿着海岸线往东北走，形成了百越民族。在距今两万年的时候，O3型中分化出了O3a3b型的一支，即是苗瑶语族的祖先。而O3型其他的人，继续向北走，成为汉、羌、藏等人的共同祖先。耶鲁大学李辉供图

褚嘉祐与他的合作者金力有一个共同之处：他们都是从医学研究开始进入分子人类学领域的。他们发现，同是病人，不同的民族或族群对同一种病的易感程度和抵抗能力都不同，这样就开始了对不同人群的基因的研究。“开始，我和金力一起，用中国28个群体的样本，看是不是有中国人自己单独的而非洲人没有的基因，结果没发现。我们也希望找到中国本地的直立人与非洲来的现代人基因融合的证据，但是也找不到。我们的‘非洲起源说’是这样来的。”

2001年，金力等人的一篇论文《东亚现代人的非洲起源：12000个Y染色体的故事》在《科学》上发表，为现代中国人的“非洲起源说”又添了一枚重磅炸弹。具体说，褚嘉祐等的第一篇论文，是应用“微卫星”标记（STR）来研究基因的，而金力等的那篇论文，是以父系遗传Y染色体标记来做研究的。金力说：“在做2001年的那篇论文时，我心想我是中国人，祖先是来自非洲来的？我非要找出一个人，不是非洲人的后代。做了覆盖整个东亚，东南亚，西伯利亚……的12000人、163个

群体以后，我都傻眼了。我都不信这个结果。我对学生说，你把数据拿来，看一下，看上去比较模糊的，你给我重新做一遍。又有400多个样本重做了，结果，还是这样子，那没办法，我只能说，大概就是这样是对的。”

这两篇论文及其相关的研究，分别在2005年和2007年获得了国家自然科学二等奖。

2005年，金力和他的复旦大学现代人类学研究中心加入了由美国国家地理协会和IBM联手赞助的“基因地理计划”。“人群迁徙是个全球范围的项目。因为迁徙的人群，不仅跨国界，还要跨大陆的。这个领域的研究已经有很多人在做，工具和手段也在发展中。结果导致各地各实验室用的分类方法和分子标记多有出入，许多数据不可比较。现在需要一个统一的标准。我们参加这个项目的好处之一，就是可以与各国的实验室相互很好地交换数据了。”

金力觉得，加入这个计划





▲贵州的长角苗族女孩，是沿云贵高原的西侧向北走的苗瑶语族的一支。图/华盖

以后，视野更广了，“会从全球的角度去关心我们的研究”。三年来，他们采样的地区，伸延到东南亚的老挝、柬埔寨等国，因为东南亚是东亚现代人类的摇篮。在中国境内，他们大致沿两条路线去采样：一条是东南沿海一直往北上去，直到胶东半岛，到达东北；一条是从内陆的川、藏、青往北走。许多数据尚在整理分析之中，论文会陆续发表，但中国人的迁徙路线，已经大致可以勾勒出来了。

从南向北：我们都是南方人

今年8月底，江苏淮安传出一条消息：该市开始兴建“南北地理分界线”，这个分界线的标志物呈球状，就建在穿淮安城而过的淮河大桥桥墩上。消息一传出，引起一片热议，新浪网上就该不该建这个南北分界标志举行投票，参与投票的155477人中，竟有145537人反对兴建，占93.6%，理由是：“完全是劳民伤财，有制造南北隔阂之嫌”。

其实，人们不必为中国的“南人”“北人”间的隔阂过分担忧。

按分子人类学家们的看

法，现在的中国人，无论是看起来相对个子矮、人却显得更精明的南方人，还是相对个高，性格却更粗犷的北方人，最早都是从南方来的。

这里先要讲一点遗传学上的原理。遗传学家们发现，遗传基因多样性最丰富的地方，其实也就是人类起源所在的地方。这个道理是，人的基因的一些位点，在一定时间段里，是会发生突变的。这种突变积累得越多，就说明这个人群的历史越长。当人群的分化产生时，常常只是一小群人走了出去，这样他们携带的基因遗传多样性就不如原来的人群高。科学家们发现，在东非的人群，遗传多样性积累得最高，这是现代人类起源于非洲，特别是东非的重要证据之一。多样性最低的，实际是美洲的印第安人，因为他们走得离非洲最远，实际是现代人类迁徙的末端。

同样，分子人类学家们看到，在中国南方的人群中，基因遗传的多样性，是高于中国北方的人群的。这是现代中国人的祖先从南向北扩展的证据。

在这里，分子人类学家们用Y染色体的标记，来描绘中国人的迁徙图，每一种标记，都是由于人群开始分化时，有一部分人带了这种新产生的基因位点走了，这种位点，就成了行进的人群的标记。“国际Y染



染色体命名委员会”把全世界的Y染色体分为从A到R的若干大的类型,而中国人所带的,主要是O型、D型,其中D型比较古老,属于“早亚洲人”,他们大概是至少6万年前就来到亚洲了;O型属于“晚亚洲人”,我们绝大多数中国人身上都带有这种基因标记,他们大概是三四万年前来到亚洲的。

O型染色体又可以分为O1、O2、O3等亚型,相对应的,有M122、M95、M119等基因位点。

现在,我们可以回到三万年前了,看看从基因图谱里展现出的我们的先民们的迁徙图景。

三万年前的时候,亚洲大地刚刚经历了一次冰川期。当携带着O型染色体的先民们来到这里时,他们可能已经发现,在他们之前,有一批现在被分子人类学家们称作“早亚洲人”的携带D型和C型染色体的人群,已经早在几万年前就在亚洲居住了。这些“早亚洲人”,当年可能是沿着海岸线前进,那里相对地形平坦,又方便捉鱼虾来吃。在冰川期的时候,海平面比今天低,许多今天的海岛当时都可以走过去。其中带有C型染色体的“早亚洲人”一直走到了今天的巴布亚新几内亚,甚至澳大利亚。

不过,今天的亚洲,“早亚洲人”分布得零零散散,在菲律宾和马来西亚一些岛屿上还有一些个子矮小,肤色之黑不逊色于非洲黑人的“小黑人”,他们的染色体基本是D型。日本的虾夷人也是D型,大家普遍猜测是氏人后代的白马藏族人 and 藏族人,也有相当多的比

例为D型,但是,他们的外貌与“小黑人”差距相当远,这是怎么变化过来的,至今还是未解的谜。在今天看来,亚洲大地上,经历过一次“晚亚洲人”与“早亚洲人”的竞争之战,技术和体力都占上风的“晚亚洲人”胜出,而“早亚洲人”,只在边边角角还有留存,剩下的,融入到各地的人群中。

携带O型染色体的“晚亚洲人”们,到达喜马拉雅山脉的南麓之后,慢慢扩张。这种扩张,并不像我们想象的,每天朝着一个目的地不停行进,而是一种人口的自然扩散。按正在耶鲁大学医学院遗传学系的做博士后的李辉的说法,当族群人口发展到一定规模,这一块地区的资源消耗差不多了,就要分出一支人,向另外的地方去。所以这种迁徙,并不从此地到彼地的步行速度为标准,它往往要经历上千年甚至上万年的历程。

“晚亚洲人”们分了两路走。

往南往东的一路,绕过老挝的崇山峻岭,进入了越南。在北部湾地区形成了侗傣语系祖先。这一支先人的Y染色体出现了O型的亚型,即O1和O2。O1型就是带有M119位点突变的那支。后来他们又从越南、广西方向进入中国,沿着海岸线往东北走,形成了百越民族。今天的黎族、侗族、水族、仡佬族、仡佬族、高山族、壮族、傣族等,都是他们的直系后代。这些民族在语言上、文化上有很强的一致性,所以也统称越人。那支带O1的人群,后来又继续向东北走,



▲ 白马藏族儿童,他们有D型染色体。资料图片



▲ 贾昆妇女,属古老的尼格利陀人(Negrito)中的一支,他们是现在留存很少的早亚洲人之一,其黝黑的肤色更像非洲人。贾昆人曾被更强大的马来人赶得走头无路,贾昆妇女现常与马来人通婚,因此其人口在减少直至消亡。

九州出土新石器遗址及Y染色体类型对照表

| 州名 | 区域 | 族群 | 代表文化 | Y染色体类型 |
|------|-----------|-----------------|------------------|----------|
| 梁州 | 秦陵以南至云贵南 | 孟高棉族群向藏缅族群过渡的人群 | 李家村文化 | O2,O3 |
| 雍州 | 陕甘青藏 | 藏缅族群的古羌人 | 仰韶文化 | O3a5 |
| 荆州 | 湖广 | 苗瑶族群 | 大溪文化 | O3a4 |
| 扬州 | 北越至苏南 | 百越族 | 大坩坑文化、良渚文化 | O1,O2 |
| 青兖徐州 | 苏北和山东 | 东夷族 | 青莲岗文化、大汶口文化、龙山文化 | 未知 |
| 豫州 | 河南 | 华族、青州的夷族、荆州的蛮族 | 裴李岗文化 | O3 |
| 冀州 | 河北 | 豫州、青州向幽州过渡的区域 | 磁山文化 | O1 |
| 幽州 | 燕山以北 | 通古斯族群、古西伯利亚族群 | 红山文化 | O1,O2,O3 |
| 并州 | 山西与蒙古部分地区 | 匈奴、突厥、乌拉尔族群 | 陶寺龙山文化、细石器文化 | O3a5,N |

其中有一支在距今约1万年前时，到达山东的胶东半岛，形成东夷人。

让金力非常高兴的是，经过三年的艰苦采样，他们在山东采到了非常有价值的样本，“把这条线接上了。”

晚亚洲人的另一支，是沿云贵高原的西侧向北走。他们主要是携带O3-M122型染色体的人们。在距今两万年的时候，这支人群中分化出了O3a3b型的一支，即是苗瑶语族的祖先，这支人，后来就往东去，到达湖北湖南一带。而其他的人，继续向北走。他们身上仍主要是O3的基因，实际上，他们是汉、羌、藏等人的共同祖先。

到了距今8000到6000年前的时候，有一支人又从这个O3的队伍中分化出来了，从甘肃一带进入了河套平原。这支人身上出现M117的基因位点突变，他们开始从事农耕，种小米。这支人，就是今天的汉族的祖先。

而若干没有进入河套的羌人部落，在距今约3000年的时候，又调头向南向西，与在西藏的土著居民融合，形成了今天的藏族。

在今年8月号的《科学美国

人》中文版(繁体)上，金力和李辉写了一篇文章《重建东亚人群的族谱》，其中画了一幅图，把距今五六千年前的中国，分为“梁、雍、荆、扬、青(兖、徐)、豫、冀、幽、并”九个州，分别对应在各州境内出土的新石器时期的文化遗址。(见图表)可以看出，当时在中国大地上，汉族的祖先，带O3的华族，只在河南的裴李岗文化中占有一席之地，其他文化，如在湖北的大溪文化、在浙江的良渚文化、在内蒙的红山文化，根据对文化遗址中古尸的古DNA的测定，分别是属于苗瑶族群、百越族、通古斯与古西伯利亚族群。当时的百越族和苗瑶族群，占据着南中国的大片土地，苏北和山东是东夷人

中国南北汉族与非汉族平均遗传距离数值

| | |
|-----------------|--------|
| 北方汉族人群之间 | 0.0056 |
| 南方汉族人群之间 | 0.0116 |
| 北方汉族与北方少数民族之间 | 0.0245 |
| 南方汉族与南方少数民族之间 | 0.0304 |
| 北方汉族与南方汉族之间 | 0.0517 |
| 北方少数民族与南方少数民族之间 | 0.0599 |



占着,北边是通古斯人群,而华夏族只是河南甘肃一带一支不大的人群。

由北向南:我们都成了中国人

汉族成为今天世界上的第一大民族,是最近几千年的事。这几千年的事,在中国的古籍中开始有了记载,而近几年在中国人基因图谱上的发现,又和这些记载——开始衔接。

首先的发现,是中国人中,南方人和北方人确实存在不同。在金力、褚嘉祐主编的《中华民族遗传多样性研究》一书中,讲到中国的北方汉族、南方汉族、北方少数民族和南方少数民族间的平均遗传距离。这种遗传距离在北方汉族人群间最小,其次是南方汉族人群间的。再其次,是北方汉人与北方少数民族间的、南方汉族与南方少数民族之间的。而北方汉族与南方汉族之间的遗传距离明显高于以上的距离,南方少数民族与北方少数民族间是遗传距离最大的。(见图表)

然而,在中国汉族中,无论北方人还是南方人,他

们的Y染色体实际上都是以O3型为主,而且绝大部分是O3a3c。金力与李辉、文波合写的一篇文章《汉族的遗传结构:文化传播伴随人口扩张》中指出:根据对基因样本的比较,他们发现,北方汉族和南方汉族中,带有O3-M122基因突变的,北方人平均是53.8%,南方人平均是54.2%,但是,在南方少数民族中普遍出现的O1-M119和O2-M95,在南方汉族人中出现频率(平均19%)要比北方汉族人的(平均5%)要高。

这是说的Y染色体,即父系遗传的情况。如果看线粒体,即母系遗传的情况,则南北方汉族有很大不同。“东亚北部的主要单倍体在北方汉族中的频率(平均55%)比在南方汉族中(平均36%)要高得多。”如果不用专业术语,通俗地说,在母系遗传中南方汉族中的“南方成分”比北方人更多。

其实这里说出的历史史实,我们是可想象的,而且也得到很多历史记录的佐证了。

在迄今的两三千前中,发源于河套地区以种植小米开始进入农耕社会的汉族,是一个从西北向东



如图可见,在图中斜线部分,人群迁徙分成了两支,一支从甘肃一带进入了河套平原,时间大约是距今8000到6000年前,他们开始从事农耕,种小米,就是今天的汉族的祖先。而没有进入河套的羌人部落,在距今约3000年的时候,又调头向南向西,与在西藏的土著居民融合,形成了今天的藏族。缅、彝、土家族等,则向今四川、云南一带扩散。耶鲁大学李辉供图



▲ 汉族、羌族和藏族人拥有共同祖先，他们都拥有O3基因。

向南的扩张过程。在这个过程中，携带O3-M122的汉族，人口大量增长，从北向南向着整个中国大陆蔓延开来，与南方种水稻的O1-M119交融。而在这个过程中，两千多年间，又经历过几次因战争和饥荒带来的人口大迁徙，大规模的有三次：即发生在晋时的“永嘉之乱”，那次的迁徙，到了南朝刘宋初年时，南渡人口已达到90多万，占当时南方总人口的1/6；第二次，是发生在唐末的“安史之乱”，使南方人口第一次超过了北方地区；第三次，是北宋的“靖康之变”，随着北方大量劳动力和先进垦殖技术的南迁，原先的“蛮荒之地”大都变成了“鱼米之乡”，在风俗习惯上南北方互相融合，经济上南强于北的局面完全确立。

人们判断，当时南迁的中原汉人，男性多于女性。这个民族迁徙与融合过程中，有相当多的汉族男性娶了土著的南方女性为妻。用李辉的简洁说法：在南方混合进南方少数民族成分最多的两个地区是浙江和

广西。浙江汉人混入本土越人的成分大约1/4-1/3。广西某些汉族群体则更高。

这里值得注意的是今天生活在中国南方的若干著名的汉族人群，如客家人、潮汕人。

2003年，李辉、潘悟云等人曾对客家人做过基因上的测试，写成《客家人起源的遗传学分析》，指出客家人与中原汉族最近，基因结构中汉族结构占80.2%，类畲族结构13%，类侗族结构6.8%。人们一直在说，客家人是从北方中原移居过来的，这从基因上得到了证实。当然，这个迁徙不是一次完成，中间又渗进了不少当时地处湖南湖北的苗瑶语族人的血液。

那么潮汕人呢？最近也有研究论文出来。由许丽娜、冯国英等人写的《常染色体STR的证据支持潮汕人的中原汉族起源》一文就是讲这件事的。文章说，潮汕人与闽南人亲缘关系最为接近。而今天福建的汉族，“在父系方面几乎完全来自北方，母系方面明确来自南方的成分也很少（大部分是不明确来源）。广府人

国家历史 淘宝网店开幕

黑色的眼睛，发现彩色的历史

<http://shop35822059.taobao.com>

一起来分享历史的宝藏！

父系方面80%以上来自北方,母系中也不低。所以全国汉族的遗传一致性是相当明显的。”李辉说。而更早在福建居住的百越族群,在今人的基因中完全看不到了,这恐怕是汉代的强制移民的结果。这让我们看到了当年的民族融合的另一面。

这里多谈一点南方汉族的语言。人们都觉得福建汉族的方言最难懂,广东人的粤语对许多人来说也如同一门外语。而李辉说:福建的方言来自于太行山两侧的上古汉语方言,又在福建孤立发展了这两千年。广东话主要来自中原西部的中古汉语,可能更接近长安音。“这些都是汉语内部的流变。语言的分化和遗传的分化是不同的过程。”

为了更好参与“寻找人类起源计划”,复旦大学现代人类学研究中心团队成员每人都测试了自己的基因。李士林是纯粹的北方汉人,携带O3-M122,不过,他母亲是满族人,祖籍浙江的金力发现自己带了古越人的M119,而原以为是土生土长上海人的李辉,带的是一个很罕见的类型,M164,属于侗傣语族,只在老挝发现过。

今天走在上海的街头,熙熙攘攘中你可以听到各

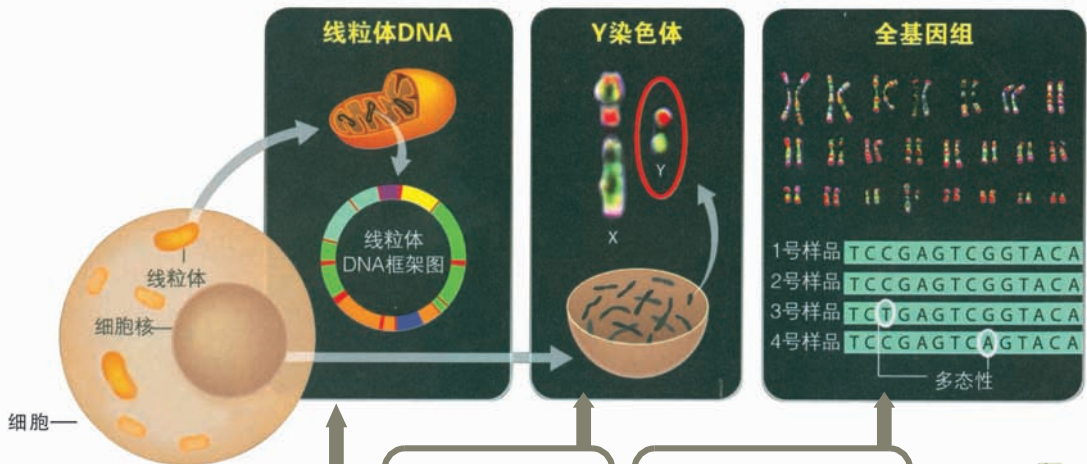
地的口音,说上海是个移民城市,一点不错。只不过,有谁会从李辉脸上那两条弯下的眉穹中,看出他遥远的与云南傣族同源的血缘?

宁夏人谭婧泽已经适应了上海的生活,虽然她的普通话中还略带一点西北口音。当年潮汕人迁居到南方时,要克服相当大的不适应,“北方来的中原汉族人群在迁到南方定居以后,在人群基因库对于环境的巨大差异的适应过程中,会产生各种代谢的不协调。在潮汕地区某些疾病的高发可能就是这种不协调的结果……”许丽娜的文章中有这样的话。不过,生活在现代社会的谭婧泽适应得很快,“我现在也不太吃很辣的东西了。要听懂上海话,半年时间就够,当然现在讲上海话还不行。不过,在复旦,大家通常也都不讲上海话,都讲普通话”。

实际上,在今天的上海,川菜、湘菜相当流行,上海人也不热衷吃“本帮菜”,倒开始吃辣的了呢。

我们从哪里来?这问题千百年来一直在搅动着我们的好奇心与求知欲。而我们往哪里去,似乎已经看得清楚。用金力的话说:“再过二百年,我们肯定都混得都一样了。”

线粒体、Y染色体和全基因组



线粒体:是一个闭合环状的、存在于细胞核外的遗传大分子,在每个细胞中存在成百上千个拷贝。因它是母系遗传,又缺乏重组,所以可以使得人们能够容易从中追踪人群的史前历史,推测人群间的母系基因交流和人群分化。它的高进化速率(约是核基因组的10-20倍)也可使人们更容易推测人群扩张、迁徙的大致时间。

Y染色体:它存在于人类男性的第23对染色体,“性染色体”中。女性的性染色体,由“XX”两条染色体组成,而男性染色体,则由“XY”两条染色体组成。Y染色体它有不重组干扰、群体内变化低而群体间变化高、遗传较为稳定等特点,适合于追溯早期人类群体起源,推测古老群体进化事件对现存人类群体多样性的贡献。

全基因组:一个生物体的基因组是指包含在该生物体的DNA(部分病毒是RNA)中的全部遗传信息。基因组包括基因和非编码DNA(即所谓“垃圾DNA”)。而对于像人类这样的脊椎动物,基因组通常指的只是染色体DNA。因此,尽管人类线粒体里包含了基因,但这些基因并不作为基因组的一部分。事实上,有时候称线粒体拥有自己的基因组,通常叫做线粒体基因组。

