

上下五千年,從南到北,中國人群祖先的基因如何流動?



少,很可能跟新時期時代後期北方人基因的擴散有關。研究人員拿新石器時代的基因和現代人的基因,分別與最古老的北方樣本(來自山東扁扁洞遺址,距今約9500年)和南方樣本(來自福建奇和洞遺址,距今約8400年)進行比對。結果顯示,新石器時代北方人群的基因與扁扁洞的樣本明顯更相似,南方人群則與奇和洞的樣本更接近;跟上文提到的那樣,兩者能互相區別。

我們從哪兒來?我們的祖先是誰?這是很多人關心的問題。研究古人類學的學者也想知道:東亞地區作為各種文化和族群的大熔爐,這些族群是如何形成的?他們之間的親緣關係如何?

雖然考古發掘的材料能為我們提供一些信息,但遺憾的是,由於缺少古DNA材料,我們對東亞人的“基因歷史”知之甚少。

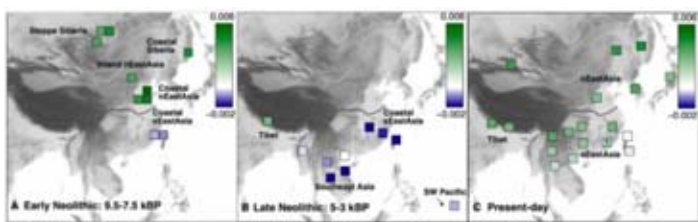
近日,中國科學院古脊椎動物與古人類研究所(以下簡稱“古脊椎所”)付巧妹團隊和同事發表在《科學》(Science)雜誌的一項研究填補了這一空白。研究人員通過對約9500年前到3000年前、生活在東亞大陸南、北方和台灣海峽島嶼的26個古人類樣本進行基因組測序,並對比已有的古DNA和現代人DNA庫,對東亞人的基因構成進行多方位研究,為我們描繪出東亞人群的基因歷史。



福建奇和洞約8400年前的2號個體頭骨

而在用現代人的基因進行比對時,研究人員驚訝地發現,現代人的基因並不會像新石器時代的那樣呈現明顯的南北區分,而是所有現代東亞人,都更接近北方扁扁洞樣本的基因,都與新石器時代的北方人有一定的親緣關係。

新石器時代早期(A)、晚期(B),以及現代人(C)與南北最古老樣本的比對結果;綠色越深代表與新時期時代北方人更接近;藍色越深則代表與新時期時代南方人更接近。現代人的比對結果(C)基本上呈現綠色。



樣本的地理位置分布(A)和測定的年代(B) 1 新石器時代南、北方已有基因差異,且比現代人的差異更大

研究人員通過對不同樣本人群的基因進行主成分分析,發現新石器時代早期南方沿海人群和北方沿海人群的基因數據會各自聚類,能互相區分。也就是說,早在9500年前,在基因層面,就已經有北方人和南方人之分。

更有趣的是,研究人員緊接着對新時期時代南北方人群和現代南北方人群的基因分別進行差異性統計,發現新石器時代南北方的基因差異,比現代南北方人的基因差異要大。這是為什麼呢?

2 新石器時代後期北方基因的傳播 帶着這個問題,研究人員進行更多的基因比對,發現當今東亞人的基因差異程度減

源成分(約22%到55%);而在中國北方漢族和東亞別的一些北方族群的基因里,也發現了南方的祖源成分(前者南方祖源成分約36%到41%,後者約35%到36%)。

雖然無法精確到具體的年份,但研究人員估計,這樣的基因流動,很可能在新石器時代晚期之前就已經開始。

3 南島語族的中國南方起源 除了揭示東亞人基因的變遷,這項研究還有一個重要貢獻:為南島語族的中國南方起源假說提供了新證據。

南島語族是指使用南島語系的族群,其分布涵蓋從台灣到太平洋西南部的小島,再到馬達加斯加的區域。學界對南島語族起源仍無定論,存在各種假說,比如“台灣起源說”、“太平洋群島起源說”、“東南亞群島起源說”、“中南半島起源說”,以及“華南起源說”等等。不同的假說都有各自的理論支持,但也存在許多爭議。

其中,“華南起源說”認為,南島語族形成於中國大陸南方地區,再擴散到太平洋各島嶼。目前,已有一些考古出土的文物和線粒體DNA的研究支持這個觀點。

南島語系分布圖 這次的古DNA研究,進一步支持了“華南起源說”。

研究人員發現,新石器時代東亞大陸沿海南方人和台灣海峽島嶼亮島人的樣本基因(距今約8400到4000年之間),與現代南島語族的基因有非常明顯的親緣聯繫。通過進一步比對來自太平洋小島瓦努阿圖的、3000年前的樣本,更發現三者基因相似度非常高,有着共同的祖源。也就是說,南島語族很可能從中國大陸南方沿海地區起源,再進一步擴散出去。

中科院古脊椎所古DNA超淨室|圖片來源:中國科學院古脊椎動物與古人類研究所徐欣

除以上提到的發現外,論文還展示了更多東亞人群基因變遷的具體情況,比如新時期沿海地區人群持續的遷徙和基因流動等等。

4 問題系統設計,求證步步小心 與歐洲相比,東亞地區雖然有大量的古人類樣本出土,但由於缺乏基因層面的研究,東亞人口的流動、不同人群如何相互融合等問題,很多時候只能通過一些文化的遺址來判斷。2010年,古脊椎動物所成立古DNA實驗室。當時還在德國讀博的付巧妹,希望有一天能夠從基因層面搞清楚這些問題。2012年,付巧妹團隊拿到了幾個關鍵的個體樣本,包括8000多年前的台灣海峽亮島的個體樣本和8000多年前的山東博山的個體樣本。雖然當時已經就南北的DNA進行比較,結果也說明一定的問題,但付巧妹認為,亮島的樣本並不能完全代表南方人群,因為島嶼的樣本很特殊,很可能由於隔離的原因與內陸的人群不同。因此,從2014年開始,她的團隊一直在尋找能夠代表南方內陸的材料,做了非常多南方個體的實驗。直到2018年底,她的團隊才做來福建奇和洞人的基因組,也就得到了真的能夠代表內陸的數據。付巧妹告訴《知識分子》,這項研究工作經歷了8年探索,最終得以發表,得益於他們對於東亞地區人群基因流動問題的整體系統性設計,另外就是找到了多個對該項目感興趣的團隊合作。查閱這篇論文的合作單位,可以看到除了古脊椎所,還有福建博物院考古研究所、平潭國際南島語族研究院、福建龍岩及漳平市博物館、山東考古與文物研究所、吉林大學、山東大學、山東省博物館、北京大學、內蒙古文物與考古研究所、西北大學等國內多家機構的名字。“這次跟相關的考古老師合作,大家都非常開心,因為有些問題也是他們一直關注的,比方南島

語族研究院的范雪春老師,他就非常非常高興,他說只要能夠做出來,從我們的研究中能夠得出一些比較重要的答案,對他來說很重要。”付巧妹說。這項研究不僅揭示出新石器時期東亞大陸南、北方人的遺傳聯繫,還為南島語族的中國南方起源假說提供了關鍵證據,為未來關於東亞人群變遷的研究打下了基礎。對於東亞早期人口流動與變遷,付巧妹認為,畫面應該會更複雜,但現在還不好說。“雖然我們看到新石器時期以來,北方人群的影響會更大,農業跟新石器人群的聯繫是非常有意義,但早期是什麼樣的不知道。在歷史時期,哪個文明的強盛可能會影響更大,但是在舊石器,我覺得可能環境會很影響變遷,總之相關研究會很有意思。我也非常非常期望能夠做出一些更早人群的工作。”付巧妹說。



族研究院的范雪春老師,他就非常非常高興,他說只要能夠做出來,從我們的研究中能夠得出一些比較重要的答案,對他來說很重要。”付巧妹說。這項研究不僅揭示出新石器時期東亞大陸南、北方人的遺傳聯繫,還為南島語族的中國南方起源假說提供了關鍵證據,為未來關於東亞人群變遷的研究打下了基礎。對於東亞早期人口流動與變遷,付巧妹認為,畫面應該會更複雜,但現在還不好說。“雖然我們看到新石器時期以來,北方人群的影響會更大,農業跟新石器人群的聯繫是非常有意義,但早期是什麼樣的不知道。在歷史時期,哪個文明的強盛可能會影響更大,但是在舊石器,我覺得可能環境會很影響變遷,總之相關研究會很有意思。我也非常非常期望能夠做出一些更早人群的工作。”付巧妹說。



人類基因組計劃 20 年後,我們開始覬覦“上帝”的角色

1986年3月,美國能源部(DOE)在新墨西哥州的聖塔菲(Santa Fe)召開會議,討論人類基因組測序計劃。諾貝爾生理或醫學獎獲得者羅納托·杜爾貝科(Renato Dulbecco)在《科學》雜誌發表文章指出,對人類基因組的測序將成為腫瘤研究的關鍵轉折點。

同年6月,加州理工學院萊諾·伊·胡德(Leroy E. Hood,1987年的拉斯克醫學獎獲得者)和勞埃德·史密斯(Lloyd Smith)改進了程序繁瑣的Sanger測序法,發明瞭世界上第一台DNA自動測序儀。

可以說,不管是從現實需要還是技術支持,對人類基因組進行測序的時機已經相對成熟。1990年4月,美國國立衛生研究院(NIH)和能源部共同發佈人類基因組5年計劃,該計劃包括人類及模式生物基因組的測序和圖譜繪製、數據收集和分析支持(算法改進、軟件設計開發等)、技術研發和轉讓等。

同年8月,NIH開始對四種模式生物進行基因組大規模測序,包括山羊支原體(Mycoplasma capricolum)、大腸桿菌(Escherichia coli)、秀麗隱杆線蟲(Caenorhabditis elegans)和釀酒酵母(Saccharomyces cerevisiae)。

兩個月後,NIH和能源部“對了對時鐘”,正式宣佈10月1日為“人類基因組計劃”官方正式起始時間。自此,先後由美國、英國、法國、德國、日本和我國科學家共同參與、被譽為生命科學“登月計劃”的“人類基因組計劃”正式踏上征程,這一走就是將近13年。

當然,這期間也出現一些“插曲”,例如生物學家、“科學狂人”克雷格·文特爾(John Craig Venter)因為與“人類基因組計劃”團隊在基因組測序數據分享政策方面存在分歧,一氣之下離開了NIH,並於1998年5月成立賽萊拉(Celera)生物技術公司,宣稱將在3年內完成人類基因組測序工作。於是NIH不得不提高效率,加快進度,以免被文特爾搶先。實際上,這種“競爭”也推動了“人類基因組計劃”更加快捷地進行。

最終的結果還是皆大歡喜的(注:在很多人看來,賽萊拉贏得了這場競賽)。國際人類基因組測序聯盟和賽萊拉生物技術公司分別於2001年2月15日和2001年2月16日在《自然》和《科學》雜誌發表人類基因組草圖。

克林頓宣佈人類基因草圖完成 “人類基因組計劃”最終完成是在2003年4月,宣告“後基因組時代”到來。那麼,“人類基因組計劃”對當今科學研究有哪些影響?

美國西北大學網絡科學研究所亞歷山大·蓋茨(Alexander J. Gates)等人在2月11日《自然》雜誌撰文指出,“人類基因組計劃”大大提昇了我們對基因及其調控元件的認識,並協助研究人員確定眾多藥物的靶點[1]。

“人類基因組計劃”最顯著的特點就是數據量龐大,項目實施的13年中,得到“注釋”的基因數量迅速增加。自2001年開始,每年關於蛋白編碼基因的學術論文數量在10000到20000篇,主要集中在CD4、TP53、TNF和EGFR等明星基因上。這都得益於“人類基因組計劃”將數據獲取方式由“單鈎釣魚”改為

“細網撈魚”。近30年,數個人類“明星基因”論文發表趨勢,從圖中可看到HPG計劃成功後,論文發表數量猛增

“人類基因組計劃”還進一步確定了基因組中非編碼序列的生物重要性。這些序列的改變不會影響蛋白的序列,但會干擾蛋白表達的網絡,進而影響生物學功能。

2000年之後,科學家對基因非編碼區域的瞭解興趣一直未減

在2001年之前,明確某種藥物全部蛋白靶點的概率不到50%。“人類基因組計劃”完成後,美國每年通過的藥物幾乎都有清楚的作用靶點說明。

“人類基因組計劃”不僅促進了生物學和生物醫學的發展,而且積極深化了多學科合作的“大科學”融合。

自本世紀初,信息科學和計算機科學的發展更加迅速,“人類基因組計劃”就是遺傳學、生物化學、分子生物學和信息科學深度合作的成功案例,此後各類“組

學”出現,共同構建生命科學的“大數據”時代。與此同時,隨着計算機算法的不斷更新迭代、測序成本的逐漸降低,個人基因組學正在走進人們的日常生活。基因組測序的大眾化更有助於我們瞭解人類自身遺傳的多樣性,最終服務醫學發展,保障人類健康。

20年後,當我們回顧人類基因組計劃時,人類已不再滿足發現基因,而是修飾基因、編輯基因,甚至是創造一個此前從未出現過的基因。下一個20年,我們還能對這些人類基本的遺傳單元做些什麼呢?至少我們已感受到了當下基因治療正如日中天發展着,我們正在收穫“人類基因組計劃”當年播撒的果實。

作者簡介:計永勝,中國科學技術大學生命科學與醫學部青年講師,曾擔任安徽醫科大學副教授,威斯康星醫學院癌症中心做博士後。

20年前,我們對基因的瞭解還很有限,以至於人類基因組草圖公佈的時候,引起了全球的矚目,從電視到報紙,媒體在連篇累牘地報道該事件。20年之後的今天,我們不再滿足於發現新的基因,而是編輯或改寫基因,瞭解人的生老病死,改良作物品種,我們甚至還創造出了此前從未出現過的基因。從某種程度上講,我們在扮演造物主的角色。

本文就帶你回顧當年激動人心的人類基因組計劃,它是如何從一個雄心勃勃的開端,走到躊躇不前、困難重重的中間,最後還是取得了一個皆大歡喜的結果。下一個20年,我們能對基因做些什麼,會有哪些驚喜的發現?

2021年2月,《科學》(2月5日)和《自然》(2月11日)雜誌出版人類基因組計劃專刊,以紀念人類基因組草圖發表20周年。

1953年,克里克和沃森發現DNA的雙螺旋結構;1977年,沃特·吉爾伯特(Walter Gilbert)和弗雷德里克·桑格(Frederick Sanger)發明瞭DNA測序的方法。

基因組學研究里程碑事件,圖片來自lehigh.edu

隨後,包括EBV病毒在內的一些簡單生物的DNA序列被測序成功。雖然有一些科學家建議對人類基因組進行測序,但面對人類基因組30億碱基對的龐大規模,持續性的研究經費成為擺在研究人員面前的難題。

