UPPSALA
UNIVERSITET

Genetiska metoder för att kartlägga den anatomiskt moderna människans expansion

Linnéa Sandell

Independent Project in Biology
Självständigt arbete i biologi, 15 hp, vårterminen 2013
Institutionen för biologisk grundutbildning, Uppsala universitet

Sammandrag

Genetiska metoder för att kartlägga den anatomiskt moderna människans ursprung, expansion och migrationsmönster är i snabb utveckling. Syftet med denna uppsats är att ge en övergripande bild av de genetiska metoder som används i kartläggningen. Vilket syfte fyller de olika metoderna, och vilka frågor återstår att besvara? Uppsatsen är en litteraturstudie av artiklar, en lärobok samt en populärvetenskaplig sammanfattning som alla publicerats under de senaste tio åren. Resultatet av studien är att:

- MitokondrieDNA och Y-kromosomsDNA används för att skapa uniparental fylogeni (på mödernet respektive fadernet) och autosomalt DNA för att fastställa regionala skillnader. Skillnader vilka utgör markörer är enkla nukleotid polymorfismer (SNP), variabelt antal tandemrepeteringar (VNTR) och strukturella variationer.
- Den grundläggande modellen som ligger till grund för slutsatser kring mänsklig migration är en seriell grundarmodell som beskrivs antingen med hjälp av kopplingsojämvikt (LD) eller Wrights ö-modell.
- De metoder som används för att analysera och strukturera genetiska data är fylogeni och demografi. Fylogenetiska studier görs utifrån genetisk distans eller icke-rekombinanta haplotyper. Demografiska studier bygger på mikrosatellitvarians eller heterozygotigrad.

Min slutsats är att det finns två huvudsakliga angreppssätt för de genetiska metoderna. Antingen behandlas neutrala genetiska skillnader, det vill säga de som är opåverkade av selektion. Alternativet är att analysera selekterade gener och söka efter anpassningar som skiljer olika folkgrupper åt, och att utefter detta dra upp teorier kring när dessa anpassningar kan ha uppstått och orsakerna därtill. Bägge angreppssätten analyserar skillnaderna statistiskt för att bygga upp matematiska modeller över troliga migrationer. Forskningsfältet är i ständig utveckling. Nya teknologier för att ta fram DNA från arkeologiska kvarlevor och mikrobiom[1] ger löften om nya upptäckter. Ökad förståelse för mönstren för transkriptionsfaktorer och proteinuttryck är också ett viktigt utvecklingsområde.

Inledning

Eftersom vi enbart ser vår historia retrospektivt kommer vi aldrig kunna uttala oss definitivt om orsakerna till människans expansion. Hur spridningen skedde har vi större möjligheter att finna svar på. Många fält, från paleontologi och klimatologi till lingvistik, har förenats i sökandet efter svaren kring människans spridning och ursprung. Genetiken är ett relativt nytt tillvägagångssätt, med sin början under det senaste århundradet. För att kunna kartlägga människans spridning och härkomst med genetiska metoder har några teoretiska och teknologiska framgångar varit av särskild betydelse.

Robert W. Holley, Har Gobind Khorana och Marshall W. Nirenbergs arbete med att beskriva den genetiska koden i nukleotider belönades med nobelpriset 1968 (Malmquist 2013). Av störst betydelse för att kunna avläsa skillnader i denna genetiska kod mellan individer och populationer har troligtvis sekvenseringsteknologin varit. Denna teknologi, som tillåter forskare att studera ordningsföljden av nukleotider i DNA:t, uppfanns under 1970-talet men har sedan 2000-talet början ersatts av nyare och snabbare metoder (Pettersson 2013). De typer av DNA som analyseras och de skillnader i detta DNA som studeras, kallade markörer, beskriver jag under avsnittet Genetisk variation.

Det fanns i början av 1900-talet en tro att den moderna genetiken omkullkastat evolutionsteorin (Donner 2003). Theodosius Dobzhansky lade med sin bok *Genetics and the Origin of Species* (1937) grunden för den moderna syntesen inom biologi, vilken förenar evolutionsteorin och

1 All genetisk information från ett prov med mikroorganismer.

populationsgenetiken, och kom att behandla människans utveckling och diversitet utifrån populationsbegreppet i motsats till det tidigare rasbegreppet (Ovesen 2013). Den moderna syntesen har gjort det möjligt att dra historiska slutsatser från den genetiska diversitet som observeras i populationer idag, bland annat de modeller jag tar upp under avsnittet Teoretiska modeller för migration.

Cavalli-Sforza och Anthony Edwards applicerade 1964 fylogenetisk parsimoni på studier av människans genetiska diversitet. Wells (2003) visar hur de försökte förklara de genetiska skillnaderna som uppmäts så enkelt som möjligt. Utifrån detta beräknade sedan Cavalli-Sforza och Walter Bodmer tiden när de mänskliga grupperna delades. Aronson *et al*. (2001) beskriver hur Emile Zuckerkandl och Linus Pauling gemensamt arbetade fram modellen för molekylära klockor, en metod för att tidsbestämma delningen mellan arter eller populationer baserat på den molekylära variationen. De presenterade idéen i *Evolutionary Divergence and Convergence of Proteins* 1965. Ovanstående två metoder är tidiga exempel på hur man mäter genetisk distans, vilket är ett angreppssätt för att strukturera genetisk information vilket beskrivs i avsnittet Struktur med fylogeni.

Det andra angreppssättet för att studera mänsklig genetisk diversitet skildrar jag i avsnittet Demografisk analys, som delas upp i mikrosatellitvarians och heterozygotigrad. Den senare metoden har sin grund i den erkända Hardy-Weinberg jämvikten. Det var i början av 1900-talet som Godfrey Hardy och Wilhelm Weinberg var och en för sig drog slutsatsen att slumpmässig parning resulterar i en jämvikt av genotypfrekvenserna i en population. Genom att studera avvikelser från denna jämvikt kan vi dra slutsatser kring tidigare demografiska händelser i populationen.

För att synliggöra resultaten av de genetiska metoderna presenteras de ofta geografiskt. Jag ägnar avsnittet Geografisk representation till att visa på exempel på hur detta kan göras. Avslutningsvis ämnar jag diskutera fram- och motgångar i forskningsfältet, samt redogöra för svaren på mina frågeställningar, vilka presenteras nedan.

Syftet med denna uppsats är att redovisa för några av de genetiska metoder som används i kartläggningen av den anatomiskt moderna människans expansion. Jag ämnar svara på följande frågeställningar:

Vilka typer av DNA används, och vilka skillnader i DNA:t utgör genetiska markörer?
Vilka är de teoretiska modeller som slutsatserna kring människans migration bygger på?
Vilka huvudsakliga metoder används för att analysera och skapa struktur i genetiska data?

Genetisk variation

Genetiska studier grundar sig på DNA och allelvarianter i DNA:t kallas markörer.

DNA

Autosomalt DNA

Består av 22 kromosomer. Har högst effektiv populationsstorlek (N_e)[2]. Varje förälder har två kopior av varje autosom[3]. Används för att studera regional tillhörighet (National Geographic Society 2013).

mtDNA

Det finns tusentals kopior av mitokondrielltDNA (mtDNA) per cell, i jämförelse med endast två

2 En kvot för att jämföra den genetiska driften hos olika populationer (Jobling *et al*. 2004).
3 Kromosom som inte är könskromosom (Nationalencyklopedin 2013a).

kopior av det nukleära genomet. Ett problem med att använda mtDNA är att dess kromosom är relativt liten, vilket gör att informationsmängden är begränsad (Underhill & Kivisild 2007). Det verkar som om kvinnor har en lägre generationstid än män, det vill säga de får barn tidigare. Detta leder till att mtDNA utsätts för högre grad av genetisk drift[4] och N_e för mtDNA sjunker i jämförelse med DNA som ärvs från bägge föräldrarna eller fadern (Jobling *et al*. 2004).

Y-kromosom
Samtidigt som X-kromosomen nedärvs två gånger så ofta från kvinnan som från mannen har män ofta också högre reproduktiv varians[5]. Detta gör att N_e för Y-kromosomen är lägre än för allt annat DNA i människan. När det är väldigt hög reproduktionsvarians hos män ökar N_e för mtDNA och X-kromosom i relation till det autosomala DNA:t. Lägre N_e för Y-kromosomen kan leda till en högre mutationshastighet för män, då små skadliga mutationer[6] kan uppträda som neutrala[7] oftare än de skulle i autosomala gener. Därför kan Y-kromosomen ha en högre grad av fixering[8] av mutationer. Bildning av spermier innebär fem till sex gånger fler celldelningar jämfört med bildning av ägg vilket kan resultera i en högre grad DNA-skador och fler mutationer (Underhill & Kivisild 2007).

Mikrober
Ett nytt område för att studera mänsklig genetisk diversitet har uppkommit de senaste fem åren, nämligen metagenomik. Man utforskar då mikroorganismerna som lever på och i människan (Lewis 2013).

Exempelvis använde Caufield *et al*. (2007) bakterien *Stretococcus mutans* (orsakar karies) i ett försök att kartlägga mänsklig fylogeni. Bakterien påminner om mtDNA då den i högst utsträckning förs mellan mor och dotter. Man kan använda sig av forntida mikrobiom från arkeologiska fynd för att mäta genetiska likheter och skillnader mellan våra förfäder och oss själva. Forskningsfältet är nytt och har stor potential i den antropologiska genetiken.

Molekylära markörer
Tidigare har mycket studier gjorts på blodgrupper, protein- och enzymvariationer, vilket nu kallas klassiska markörer (Fullerton 1997). Dagens forskning utgår i stor grad från molekylära markörer, vilka beskrivs nedan.

SNP (från eng. single nucleotide polymorphisms)
Polymorfism betyder att det finns flera (poly) former (morf) av ett objekt. SNP betyder således en skillnad på en nukleotidsite i DNA:t (Griffiths *et al*. 2008). I alla människors genom, det vill säga det mänskliga genomet, finns 10.000.000 SNP (The International HapMap 3 Consortium 2010). Dessa polymorfismer kan uppstå av två olika anledningar: bassubstition eller insertion-deletion (indel). Bassubstitution sker när en kvävebas byts ut till en annan, antingen genom ett fel i inkorporeringen under replikation eller genom mutagenes (Jobling *et al*. 2004). Detta leder till en SNP då det finns fler alternativ för vilken kvävebas som finns på den specifika platsen i DNA-sekvensen jämfört med en annan individ (Wells 2003).

---

4 Slumpmässig effekt på genfrekvenserna i en begränsad population (Nationalencyklopedin 2013b).
5 Variation i antalet avkomma reproducerat av en grupp individer (Jobling *et al*. 2004)
6 Alla förändringar som producerar nya alleler kallas mutationer (Jobling *et al*. 2004).
7 Alleler vars frekvens enbart påverkas av genetisk drift
8 En mutation är fixerad när det är den enda alleltypen i en population.

Indeler skapar markörer genom att nukleotider adderas eller subtraheras. Detta gör att längden på homologa DNA-segment varierar mellan individer. Man förutsätter att det i varje generation finns en specifik sannolikhet för att en indel ska fixeras. Genom att jämföra längden på två homologa DNA-segment kan man då dra slutsatser kring hur länge de divergerat från varandra, det vill säga antalet indeler per läge som ackumulerats under divergensen av de två sekvenserna. Fördelen med indeler är att längden på två homologa segment av DNA kan divergera i mycket hög grad, samt att de sker mycket mer sällan en bassubstitution, vilket gör att sekvenser med stort genetisk distans[9] kan jämföras (Ogurtsov *et al*. 2004).

VNTR (från eng. Variable Number of Tandem Repeats)
Tandemrepeteringar[10] av samma sekvens genomgår ofta mutationer. Antalet repetitioner kan variera från ett dussin till mer än hundra stycken för olika genom. De tandemrepeteringar som varierar på detta sätt mellan individer kallas VNTR (Griffiths *et al*. 2008). Tandemrepeteringar på under tio nukleotider kallas korta tandemrepeteringar (STR från eng. short tandem repeats) eller mikrosatelliter. Exempelvis använde Eriksson *et al*. (2012) di- och trinukleotid mikrosatellit markörer för att representera den genetiska variationen i sina populationer. VNTR klassificeras efter en av tre egenskaper: storleken på den repeterade enheten, antalet enheter i stegen eller nivån av variabilitet (Jobling *et al*. 2004).

Tabell 1 visar vilka typer av markörer som passar att användas för analysen av olika DNA.

Tabell 1 Vilket DNA som används för vilka markörer.

Autosomala	Klassiska markörer
	Mikrosatelliter (VNTR)
	RFLP (eng. restriction fragment length polymorphism)[11]
	SNP
	Alu insertioner
	Beta globin
	CNP (copy number polymorphism)
mtDNA	RFLP
	HVS (eng. hypervariable segment)[12]
Y kromosom	Binära markörer[13]
	Mikrosatelliter

CNP
En strukturell polymorfism i DNA:t. Antingen en duplikation eller en deletion i kopieantalet av en gen som förekommer med högre frekvens än 1 % i populationen (Campbell *et al*. 2011). Jakobsson *et al*. (2008) jämförde genotyp-, haplotyp- och kopieantal-variation i mänskliga populationer över hela världen. CNP är alltså en relativt ny markör som också användes av The International HapMap 3 Consortium (2010). Campbell *et al*. (2011) menar att CNP som varierar stort i kopieantal mellan

9 Genetisk distans mellan två homologa DNA sekvenser definieras som det antal mutationer per nukleotidplats som blivit fixerad under evolutionens gång i sekvenserna från deras senaste gemensamma anfader (Ogurtsov *et al*. 2004)

10 Detta är en rad (tandem) av repeteringar av samma sekvens, exempelvis ACACACAC.

11 Variation i sekvensen hos DNA som påvisas genom klyvning med restriktionsenzymer (Nationalencyklopedin 2013g)

12 Del av mtDNA som visar särskilt hög variation i DNA-sekvensen (Jobling *et al*. 2004).

13 En polymorfism som enbart har två allelvarianter (Jobling *et al*. 2004)

populationer kan vara utfallet av populations-specifik selektion och användas som indikatorer på nyliga demografiska händelser.

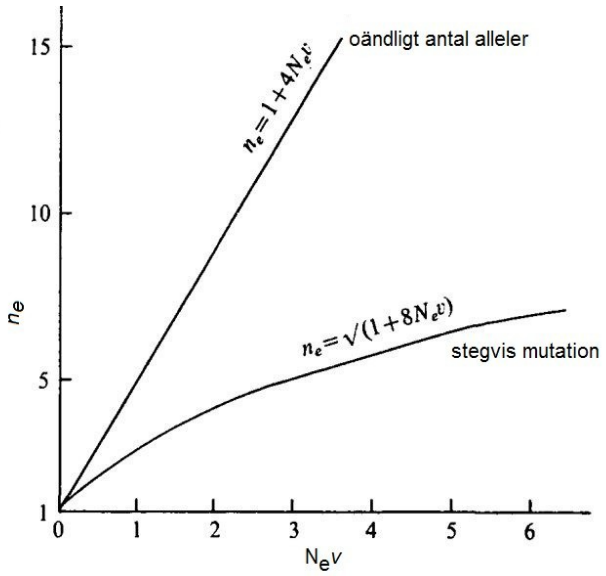
Haplotyp
Kombinationen av alleler på samma kromosomala homolog kallas haplotyp (Griffiths *et al*. 2008). Jobling *et al*. (2004) förklarar hur rekombination leder till nya kombinationer av alleler på samma DNA-molekyl vilket ökar den haplotypa diversiteten. Det internationella HapMap projektet skapar även det struktur i sina identifierade SNP genom att samla ihop de SNP som ärvs gemensamt till haplotyper (The International HapMap 3 Consortium 2010). Även Campbell *et al*. (2011) undersökte om CNP kunde kopplas till SNP haplotyper.

Mutation
Mutation är den enda kraft som skapar ny genetisk variation. Om en allel får en ny mutation kommer just denna allel automatiskt minska i frekvens i populationen. Vi kan beräkna tiden som gått från det att en allel först uppkom om vi har ett uppskattat värde för mutationshastigheten och allelens ursprungliga frekvens. Tidsuppskattningar baserat på mutationshastighet fungerar bara över väldigt långa tidsrymder, men är också av den anledningen till god användning i studier av demografi. Kimura och Crow (1964) visade hur de motstående krafterna av ökad heterozygositet och genetisk drift i en population leder till ett maximalt antal alleler för varje locus, en förväntad grad av genetisk diversitet.

Neutrala mutationer
Den allra största delen av variationen i genomet påverkar inte fenotypen (Jobling *et al*. 2004). Kimura (1970) beskrev sannolikheten för neutrala alleler att fixeras i en population. För att kunna beräkna tiden som gått från det att en gen uppkommit krävs en konstant hastighet (liknande en diffusionskoefficient). Eftersom selektion varierar i styrka och därmed påverkar hastigheten med vilken en gen ökar eller minskar i frekvens blev Kimuras teori om neutrala alleler och genetisk drift av oerhörd betydelse för de populationsgenetiska modellerna (Wells 2003).

Oändligt antal alleler
Kimura och Crow (1964) föreslog en modell där en muterad allel ses som olik alla andra alleler som tidigare funnits i populationen. Modellen fungerar relativt bra vad gällande SNP, eftersom sannolikheten för att samma nukleotidsite ska muteras är mycket liten. Vad gällande VNTR vars längd kan bli både kortare och längre, är sannolikheten att en förändring sker i samma tandemrepetitionssekvens större. Då passar en stegvis mutationsmodell bättre.

Stegvis mutationsmodell
Ohta och Kimura (1973) presenterade modellen för att illustrera evolution i mikrosatelliter. Längden varierar i en konstant hastighet oberoende av tandemrepeteringens längd – det är lika stor sannolikhet att längden på de repeterande enheterna kommer att minska som att den kommer öka. Detta gör att det effektiva antalet alleler (n_e) är lägre för stegvis mutation än det är i oändligt antal alleler-modellen, se figur 1.



Figur 1 Visar skillnaden i n_e mellan den oändligt antal alleler och stegvis mutationsmodellen. $N_e v$ står för effektiv populationsstorlek multiplicerat med mutationshastigheten per loci per generation. Omritad efter Ohta och Kimura (1973).

Teoretiska modeller för migration

Till skillnad från drift, mutation och selektion kan migration inte förändra allelfrekvenser på artnivå. Istället studerar man den genetiska variationen inom subpopulationer. Migration in till nya obebodda områden kallas kolonisering och leder till grundareffekter[14]. I populationsgenetiska modeller används migration för att beskriva genflöde mellan populationer, en kraft som motverkar genetisk differentiering (Jobling *et al*. 2004). Varje generation utgör bara en del av den genetiska sammansättningen i föräldrapopulationen. Styrkan hos den genetiska driften är beroende av populationens N_e, vilken i sin tur varierar hos varje population till följd av flaskhalsar[15], reproduktiv varians och fekunditet[16] (Jobling *et al*. 2004).

Genom att anta att drift balanseras av genflöde förmodas den genetiska diversiteten, F_{ST}[17], vara konstant över tid. Genom att analysera detta teoretiskt balanserade värde med F_{ST} värden för verkliga populationer kan vi dra slutsatser kring migrationshastigheter (Gillespie 1998, Jobling *et al*. 2004). De högsta nivåerna av modern genetisk variation förväntas finnas i den geografiska regionen för människans ursprung. Grundareffekterna reducerar den genetiska diversiteten bort från denna punkt.

Genom att undersöka spridningen av allelvarianter i och mellan populationer kan vi dra slutsatser kring nivån av struktur mellan populationerna, samt utforska ordningen och tidpunkten för deras uppkomst (Underhill & Kivisild 2007). Modeller som förutsätter att reglerna som påverkar evolutionen av alleler inom och mellan populationer inte förändras över tid kallas jämviktsmodeller (eng. equilibrium population structure models). Dessa är bra grunder för att testa hypoteser och

14 Sker under kolonisering av nya obebodda områden. En del av föräldrapopulationen ger sig av och tar då bara med sig en del av den genetiska diversiteten från föräldrapopulationen. Detta gör att den genetiska diversiteten minskar i de nya populationerna (Jobling *et al*. 2004).

15 Begreppet syftar till att en population minskar i storlek, vilket gör att bara en del av den genetiska diversiteten kvarstår (Jobling *et al*. 2004).

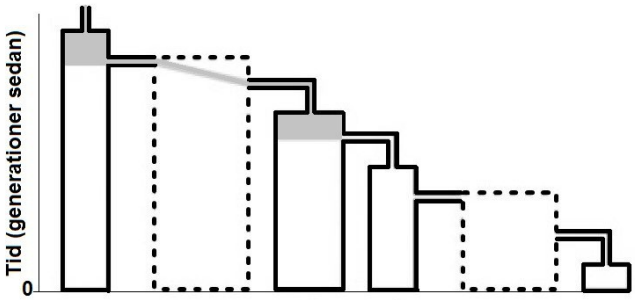
16 Potentiell fortplantningskapacitet (Nationalencyklopedin, 2013f)

17 I fallet där slumpmässig parning sker inom varje delpopulation kan man beskriva genotypfrekvenserna för hela arten med F_{ST} (Gillespie, 1998)

förutse mönster i genetisk variation. Dock är mer komplexa modeller som förutsätter att de evolutionära premisserna förändras över tid, så kallade icke-jämviktsmodeller, ofta närmare verkligheten (DeGiorgio *et al*. 2011). Nedan beskriver jag några av teorierna forskningen använder sig av.

Seriell grundarmodell

Presenterades av Ramachandran *et al*. (2005), och är en icke-jämviktsmodell. Den baseras på att koloniseringen börjar från en källa, en grundarpopulation, utifrån vilken ett antal individer regelbundet migrerar utåt och grundar en ny population. De nya populationerna innehåller bara en del av den genetiska diversiteten från föräldrapopulationen. Detta mönster upprepas till *K* populationer skapats. Man kan använda modellen baklänges, och utifrån de populationer som finns idag vandra bakåt i tiden och finna tidpunkten då populationerna måste ha delats från varandra. På detta sätt har man kunnat se en minskning i den förväntade heterozygotigraden beroende på geografisk distans från en förmodad källa, grundarpopulation, i Afrika (DeGiorgio *et al*. 2011). När en liten delgrupp av en större population etablerar sig själv som en separat och isolerad enhet uppstår grundareffekten (eng. founder effect). Eftersom delgruppens genpool bara bär en del av den genetiska diversiteten hos föräldrapopulationen ökar homozygotfrekvensen i populationen (Medical Subject Headings 2013). Figur 2 visar modellen grafiskt.



Figur 2 Illustration av seriell grundarmodell. De helstreckade kolonnerna representerar bevarade (nu levande) populationer, och de streckade kolonnerna representerar utdöda populationer. Kopplingarna mellan populationerna visar på migration innan etableringen av den nya populationen. För att ta reda på de två mittersta populationernas släktled söker man de grå fälten, vilket är koalescenspunkterna för släktleden. Omritad efter DeGiorgio *et al*. (2011)

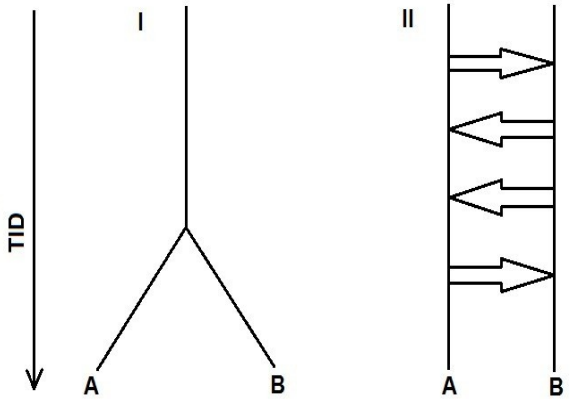
Den seriella grundareffekten kan ses antingen som en arkaisk seriell grundarmodell i vilken grundarprocessen startade långt tillbaka i tiden, benämnd Kopplingsojämvikt (LD från eng. Linkage Disequilibrium), eller som en nätaktig regionsmodell där flaskhalsar mellan kontinentala regioner är mer kraftiga än de inom de kontinentala regionerna, benämnd Wrights ö-modell (DeGiorgio *et al*. 2011). Dessa två modeller beskrivs nedan.

LD

Alleler vid närliggande loci på samma kromosom segregerar inte slumpmässigt under rekombination, utan är kopplade till varandra. Kopplade loci delar ursprung, och påverkas gemensamt av de populationsgenetiska krafterna. Om ett loci är föremål för positiv selektion sägs de närliggande ”lifta” med detta loci och når höga frekvenser, trots att de i sig själva är selektivt neutrala. Man undersöker huruvida specifika alleler vid olika loci associeras med varandra mer eller mindre ofta än vad som kan förväntas av slumpen. Analysen av denna icke-slumpmässiga koppling kallas LD (Ardlie *et al*. 2002, Jobling *et al*. 2004). Alltså, gener som befinner sig nära varandra på en DNA-sträng sägs vara kopplade till varandra eftersom de nedärvs tillsammans. Genom rekombination kan dessa kopplade gener separeras och nedärvas separat och ge uppkomst till två nya genetiska släktled. Det har också uppmäts ökad LD med ökat geografiskt avstånd från Afrika, vilket stöttar en seriell grundarmodell (Jakobsson *et al*. 2008). Detta kan ses i Figur 3I.

Ö-modell
Jobling *et al*. (2004) beskriver Wrights ö-modell som en metapopulation[18] uppdelad i lika stora delpopulationer, symboliskt kallade öar. Gillespie (1998) förklarar att man genom att förutsätta ett högt antal delningar av en population och konstant migrationshastighet *m* för alla populationer kan dra upp teorier kring en arts geografiska spridning. Genflödet mellan delpopulationerna är *m* per generation. Denna migrationshastighet är kopplad till populationens underliggande struktur F_{ST} genom: $F_{ST} = 1 / (1 + 4Nm)$.

Förutsättningen för modellen är att migration och genetisk drift är de enda faktorer som påverkar ö-populationen. Migration för in ny genetisk variation medan den genetiska driften eliminerar den genetiska variationen, då den slumpmässigt kommer att leda till att vissa gener fixeras eller elimineras i populationen. När vi vet delpopulations storlek och dess F_{ST} kan vi beräkna en teoretisk migrationshastighet (genflöde) *m* (Gillespie 1998). Migration är således av stor betydelse och inräknat i Wrights ö-modell, se figur 3II.



Figur 3 I visar på LD-modellen, där de genetiska likheterna mellan populationerna A och B kommer från en gemensam anfader. II visar på Wrights ö-modell, där de genetiska likheterna mellan populationerna A och B uppkommit från kontinuerlig migration mellan dem.

Kimura och Weiss (1964) introducerar geografisk substruktur i Wrights ö-modell. De benämner den språngbrädemodellen (eng. stepping stone model). Modellen hanterar genflöde mellan grannliggande delpopulationer och förutsätter samma migrationshastighet mellan delpopulationerna. Deras slutsats är att tendensen av slumpmässig lokal differentiering beror på antalet dimensioner; den är starkast i en dimension, exempelvis längs med en vattenlinje, och blir mindre med ett ökat antal dimensioner (Kimura & Weiss 1964).

Struktur med fylogeni

Det enklaste sättet att visualisera delat ursprung är genom att följa släktled bakåt i tiden och finna platsen där de två släktleden möts, deras senaste gemensamma anfader (Eriksson *et al*. 2012). Detta ger indirekt information om populationers rörelser.

Genetisk distans

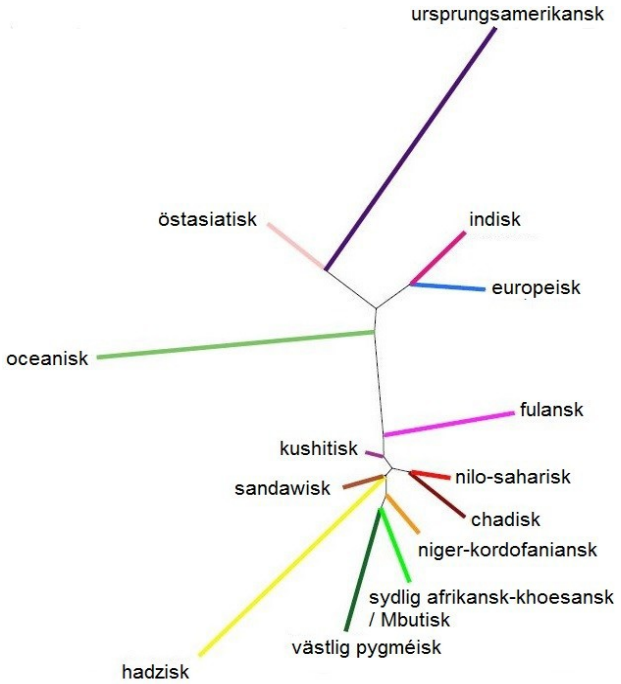
Genom att mäta genetisk distans kan man fastställa migrationshändelser och genetisk drift men dock inte uttala sig om släktskap. Genetisk distans kan mätas genom allelfrekvenser eller med komponentanalys, se nedan (Jobling *et al*. 2004).

Allelfrekvenser
Allelfrekvenser baseras oftast från bassubstitutioner. Varje nukleotidsite agerar som en oberoende

18 Flera populationer som agerar som en enda population.

timer vilket fungerar bra med korta DNA-sekvenser. Li *et al*. (2008) kunde med hjälp av SNP särskilja tydliga genetiska skillnader kopplat till geografiskt ursprung. Begränsningen med detta är att SNP bara kan ha två stadier; matcha eller icke-matcha, vilket gör att de snabbt når mättnad när den genetiska distansen ökar. Detta gör att genetisk distans bara kan bestämmas för långsamt evolverande icke-neutrala sekvenser (Ogurtsov *et al*. 2004).

Nyss har man börjat basera allelfrekvenser på individuellt särskiljbara indeler. Det är mindre risk att de mättas än substitutionsbaserad genetisk distans eftersom två homologa segment av DNA i mycket hög grad kan divergera i längd. För det andra är indeler mindre vanliga än bassubstitutioner, vilket gör att sekvenser med stort avstånd kan jämföras utan att få för många träffar (Ogurtsov *et al*. 2004). Då använder man sig av ($\delta\mu^2$), se tabell 2.



Figur 4 Visar den genetiska distansen beräknat på allelfrekvenser mellan världens folkgrupper. Omritad efter Tishkoff *et al*. (2009).

Genom att illustrera den genetiska diversiteten med en stjärnfylogeni antar man att alla populationer har haft samma tid att evolvera sedan sitt gemensamma ursprung med de övriga grenarna av trädet (figur 4).

Tabell 2 Beskrivning av tre koefficienter använda av Tishkoff *et al*. (2009) som uppskattar genetisk distans.

Koefficient	Beskrivning/Användning
D^2	Uppskattar F_{ST}. Förutsätter den neutrala teorin. Känslig för nya differentieringshändelser.
R_{ST}	Analogt för F_{ST}. Baserat på stegvis mutationsmodell. Bra för att uppskatta relationer mellan tidigt separerade och geografiskt isolerade populationer, det vill säga äldre differentieringshändelser.
($\delta\mu^2$)	Utvecklad speciellt för mikrosatellitmarkörer. Antar stegvis mutationsmodellen. Känslig för äldre differentieringshändelser.

Komponentanalys
Tillåter oss att reducera det multidimensionella rummet som utgör mänsklig genetisk diversitet till

en eller två förståeliga dimensioner (Jobling *et al*. 2004). Den vanligaste är PCA (eng. Principial Component Analysis). Cavalli-Sforza *et al*. (1994) använde denna metod när de kartlade allelfrekvensskillnader mellan världens populationer. Det är en generell metod för att representera högdimensionell data, individer eller populationer, i ett mindre antal dimensioner. PCA kan beskrivas som en algoritm som upprepande gånger söker efter ortogonala axlar[19] längs med vilka objekten visar högst varians, och ger sedan positionerna för objekten längs med de axlarna (de huvudsakliga komponenterna). Den relativa placeringen av dessa objekt längs de första huvudsakliga komponenterna ger en uppskattning av kovariansmönstren bland individer i det större datasetet och används för att utforska variationsstrukturen i provet (François *et al*. 2010).

När Francois *et al*. (2010) undersökte den genetiska variationen i Europa med PCA upptäckte de att gradienten i de huvudsakliga komponenterna var rätvinkliga till expansionsriktningen. En teori kring varför är en återkommande grundareffekt under expansionen kallas surfar-fenomenet. Idén är att det i gränsområdet för expansionsriktningen skapas områden där lågfrekventa alleler har möjlighet att öka i frekvens och till och med fixeras, medan de kan försvinna helt i den huvudsakliga expansionsriktningen (François *et al*. 2010).

Tishkoff *et al*. (2009) fann 72 huvudsakliga komponenter (PC) i sitt globala prov. PC1 stod för 19,5 % av den genetiska variationen, och särskilde alla icke-afrikaner från afrikaner. PC2 stod för 5,2 % av den globala genetiska variationen och särskilde oceanier, östasiater och ursprungsamerikaner. PC3 stod för 3,5 % av den genetiska variationen och särskilde Hadza-folket, det enda kända bevarade jägar-samlarfolket.

Icke-rekombinanta haplotyper

Ger en indikation på tiden som gått. Kallas även uniparental fylogeni, eftersom det bygger på bara ett föräldraled. För att man ska kunna följa släktled krävs att det DNA man undersöker är icke-rekombinant. Det finns två typer av DNA i människan som inte rekombinerar: mtDNA och den icke-rekombinanta delen av Y-kromosomen (NRY efter engelskan non-recombining Y-chromosome) (Oppenheimer 2012).

mtDNA

Enligt endosymbionthypotesen har mitokondrien bakteriellt ursprung. Dess cirkulära kromosom genomgår inte mitos och rekombineras därför aldrig, vilket gör att vi kan använda dess DNA för att undersöka rakt nedstigande släktled. Eftersom mtDNA enbart ärvs på mödernet, brukar den koalescenspunkt man uppnått hänvisas till som mitokondriella Eva (Wells 2003).

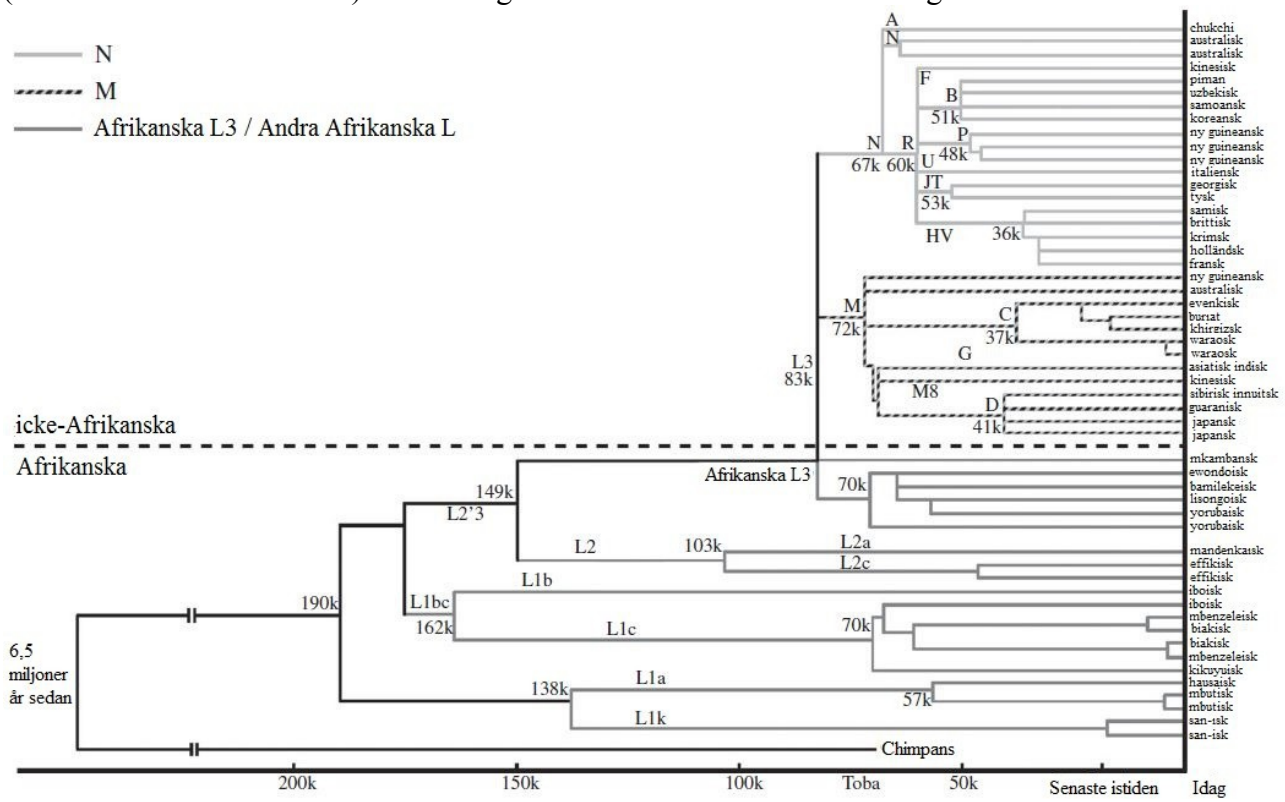
Underhill och Kivisild (2007) har gjort en omfattande sammanställning av mtDNA fylogeni. Släktträdet delar sig tidigt i sekvenser som enbart finns i Afrika och bara en haplogrupp[20], L3, som afrikanerna delar med resten av världen. Alla icke-afrikanska mtDNA släktled kommer från två avstickningar från haplogrupp L3, nämligen M och N. N-släktledet har gett upphov till dotterkladen R utanför Afrika (Underhill & Kivisild 2007).

Från dotterkladen R har två haplogrupper uppkommit i Europa: T och H. Dessa särskiljer sig genetiskt från varandra i lika hög grad som de var och en skiljer sig från haplogrupp B (Asien, Oceanien) eller haplogrupp P (Papuan, Australien). Haplogrupp B särskiljde sig tidigt från sin modergrupp R. Detta gör att man intuitivt förväntar sig att den borde vara spridd i nästan lika hög grad som haplogrupp R. Dock är den lokalt begränsad till den östra hemisfären; troligen tillkom den inom den östasiatiska grundarpopulationen (Underhill & Kivisild 2007).

19 Beskrivna som linjära kombinationer av multivariata observationer (Francois *et al*. 2010)

20 Haplogrupp är en samling av haplotyper.

Européer och Främre Orientens populationer har fått sina maternala släktträd från haplogrupp N och R. N1, W, X, JT och R0 och U; de första tre av dessa härstammar från haplogrupp N, medan de senare tre delar ursprung i R. Det finns inga signifikanta frekvensskillnader i dessa större haplogrupper mellan geografiska distinkta populationer i Europa. Omfattande sekvensering kan visa på distinkta regionala skillnader, exempelvis uppträder haplogrupp M7a bara i Japan och Sydkorea (Underhill & Kivisild 2007). De kartlagda släktleden för mtDNA ses i figur 5.



Figur 5 Mänsklighetens mitokondriella släktled. k = tusen år sedan. Omritad efter Oppenheimer (2012).

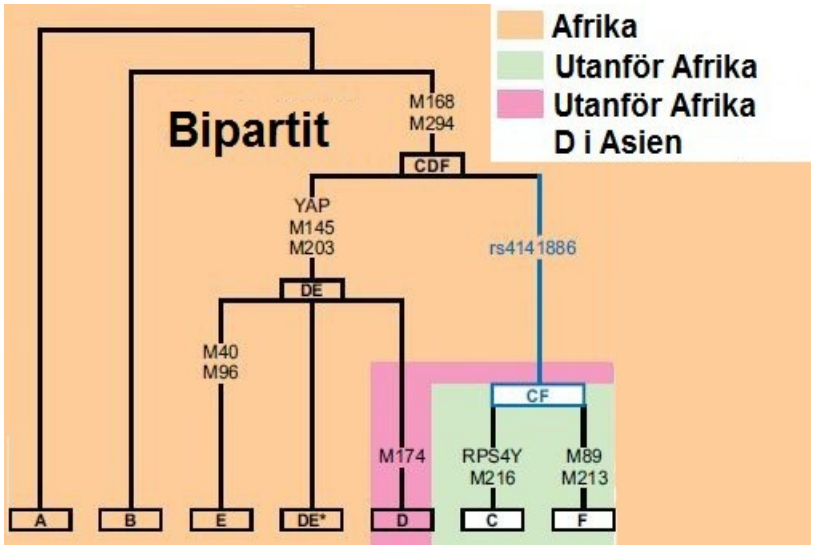
NRY

Innehåller den största delen icke-rekombinant DNA i det mänskliga genomet och kan räknas till ett av de mest informativa haplotyp-systemen (Underhill & Kivisild 2007). Första delning i NRY-släktträdet är haplogrupperna A och B, vilka bara existerar i Afrika. Dessa grupper är genetiskt diversa med många delhaplogrupper som är geografiskt särskiljbara från varandra, något som stämmer överens med en historia av populationsfragmentering, isolering och vidare återexpandering i Afrika (Underhill & Kivisild 2007).

Polymorfism M168 representerar vissa afrikanska och alla icke-Afrikanska haplogrupper (Wells 2003, Oppenheimer 2012). De kontinentala genpooler av NRY inkluderar haplogrupp DE i Afrika och Asien, haplogrupp C i Östra Asien, Oceanien och Nordamerika samt en global spridning av en annan icke-afrikansk som kännetecknas av polymorfism M89 (Underhill & Kivisild 2007). M130 är en mutation på NRY som bara finns till öster om Kaspiska havet. Den är lågfrekvent i Indien, men ökar sedan gradvis över Malaysia, Nya Guinea och finns i 60 % av Australiens aboriginska män (Wells 2003).

Alla asiatiska män delar M89 och M9. På M9 är det tre polymorfismer som skiljer de asiatiska männen: M20, M175 och M45. M45 gav upphov till dotterkladen M242 som kommit att dominera i populationer i Östasien, Sibirien och den amerikanska ursprungsbefolkningen (Wells 2003).

Haplogrupperna C, DE och F utgör majoriteten av världens Y-kromosomer (Underhill & Kivisild 2007). Det har funnits olika teorier kring delningen av haplogrupp C (Östasien, Oceanien och Nordamerika) och DE (Afrika och Asien). Genom omfattande sekvensering fann Underhill & Kivisild (2007) en mutation som delas av C och F som inte finns i de andra haplogrupperna. Detta ledde till slutsatsen att det fylogenetiska NRY-trädet följer en bipartit-modell, se figur 6.



Figur 6 Bipartit struktur av NRY-haplogrupper. Omritad efter Underhill och Kivisild (2007).

Demografisk analys

Vi använder genetisk variation, diversitet, för att uttala oss om demografiska händelser. Demografiska händelser sker på populationsnivå, och man använder därför populationsgenetiska modeller. En population kan hänvisa till en praktisk eller teoretisk enhet (Jobling *et al*. 2004).

Mikrosatellitvarians

Mikrosatellitvarians är mer känslig för mutationer som särskiljer de äldre släktleden. Rosenberg *et al*. (2002) använde sig av mikrosatellitvarians för att utröna skillnader i det mänskliga genomet. De använde sig av klusteralgoritmer och delade in de individuella proven i olika grupper. Detta angreppssätt används ofta för att skapa struktur i studier av mikrosatellitvarians. Jakobsson och Rosenberg (2007) förklarar att resultatet av enkel klusteranalys representeras som en matris, där varje individ ges en 'tillhörighetskoefficient' (eng. membership coefficient) för varje kluster. Denna tillhörighetskoefficient tolkas som sannolikheten av tillhörighet, eller som andelen av genomet som hör till klustret. Varje individs tillhörighetskoefficienter summeras till 1, över *K* antal kluster. Klusteralgoritmer bygger på matriser av C individer (som rader) över K kluster (som kolumner). Man kan på så sätt referera till uppkomsten av olika mutationer som att den tillhör nivån K = 2 exempelvis.

Med mtDNA har man kunnat särskilja att alla icke-Afrikanska populationer särskiljer sig från de Afrikanska vid K = 2, och att nästa delning över de stora kontinenterna sker först vid K = 5 (Underhill & Kivisild 2007). Detta tyder på att koloniseringen av kontinenterna skedde snabbt. Man kan föreställa sig K som en beskrivning av upplösningen på informationen. Med låg upplösning (K = 2) kan bara två grupper särskiljas, vid högre upplösning (K = 5) blir fler delpopulationer särskiljbara. Först vid K = 6 kan de centralasiatiska och ursprungsamerikanska populationerna särskiljas. För att öka upplösningen kan man även sekvensera större delar av genomet och provtesta fler individer (Underhill & Kivisild 2007).

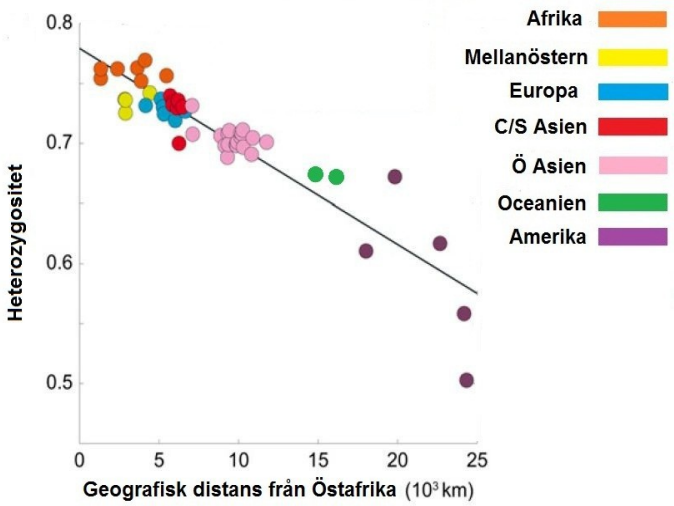
Problemet med klusteralgoritmer är att oberoende analyser av samma data kan resultera i flera distinkta slutsatser även när man använder samma grundförutsättningar (Jakobsson & Rosenberg 2007). Man diskuterar ofta Bayesiskt perspektiv – att man väljer att behålla eller förkasta hypoteser baserat på sannolikhet för det specifika utfallet.

Heterozygotgrad

Heterozygoti är förekomsten av olika alleler på motsvarande locus i de båda kromosomerna i ett kromosompar (Nationalencyklopedin 2013c). F används ofta som inavelskoefficient och står för andelen homozygoter i en population (Kimura & Crow 1964).

F_{ST} (från eng. Fixation Statistics) är ett mått på populationsstruktur. En uppskattning på hur mycket frekvensen av heterozygoter avviker från den som förmodas under Hardy-Weinberg jämvikt. F_{ST} jämför medelvärdet för den genetiska variationen hittad inom subpopulationer till den genetiska diversiteten i metapopulationen[21] (Jobling *et al*. 2004). Man använder parvisa F_{ST} för att jämföra den genetiska differentieringen hos två olika populationer (Campbell *et al*. 2011). Det finns även andra statistiska metoder för att undersöka genetisk differentiering, exempelvis V_{ST} vilken beräknar variansen i hybridiseringsvärden inom en population jämfört med variansen delad mellan populationer (Campbell *et al*. 2011).

Många arter täcker så stora geografiska områden eller har så effektiva barriärer av migration i form av exempelvis bergskedjor att de inte kan agera som en enkel, slumpmässigt parande population. Wright introducerade teorin om isolering genom distans (IBD från eng. isolation by distance) 1943, vilken kopplar geografisk distans till genetiska distanser. Den beskriver hur partnerval begränsas av geografisk distans, vilket leder till ackumulering av lokala genetiska skillnader. Ett annat sätt att beskriva det är att hindrad migration leder till inavel som leder till en ökad sannolikhet för homozygoti genom nedärvning. I dessa fall kommer genetisk differentiering uppstå i delpopulationerna. Regression av genetisk distans beräknas som $F_{ST} / (1 – F_{ST})$. François *et al*. (2010) visade hur den genetiska differentieringen ökar med geografisk distans från Afrika, se figur 7.



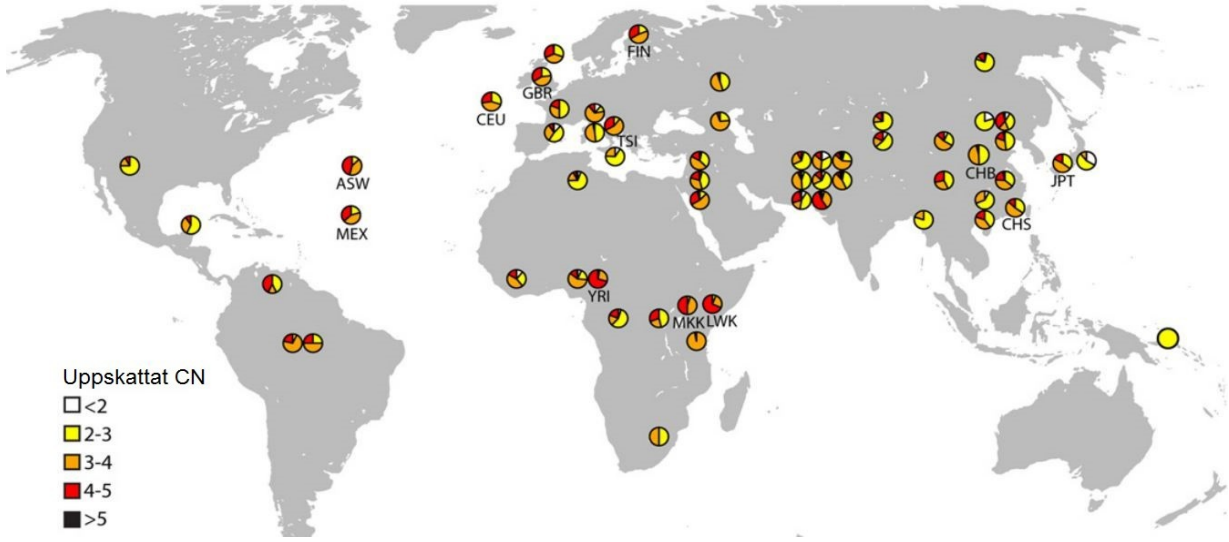
Figur 7 Heterozygositet beroende på geografisk distans från Östafrika. Varje punkt representerar en population. Omritad från DeGiorgio *et al*. (2011).

21 System av lokala populationer av en viss art vilka är delvis isolerade från varandra men mellan vilka det ändå äger rum en viss trafik av individer och utbyte av gener (Nationalencyklopedin 2013d)

Geografisk representation

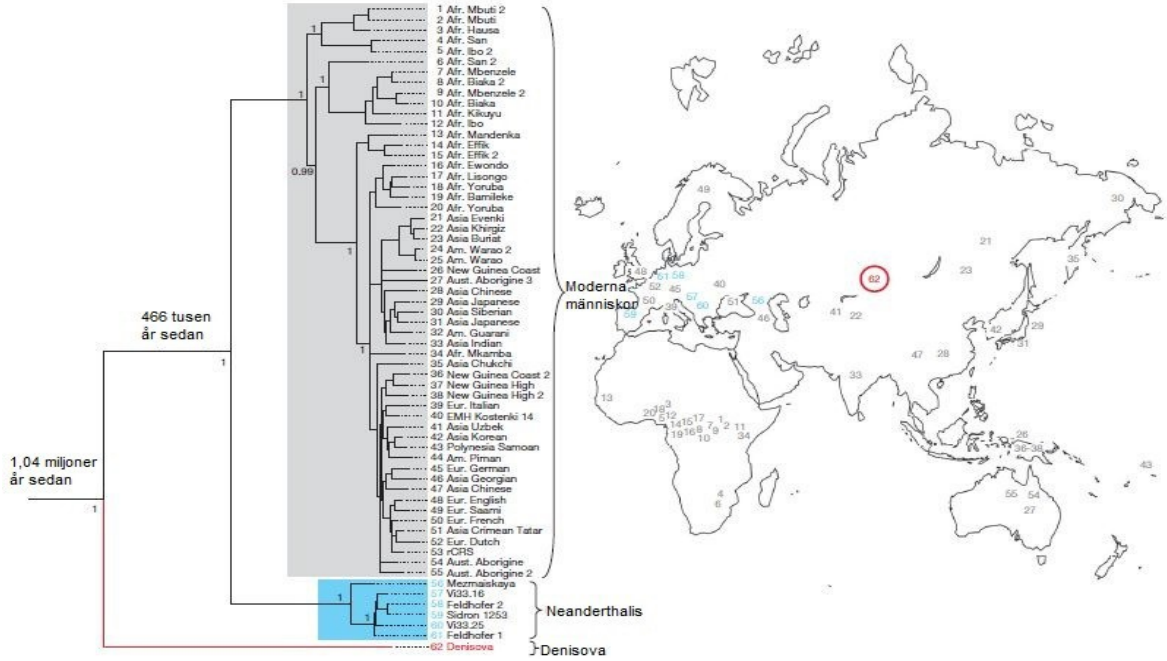
Genetisk differentiering påverkas av geografi och leder till ackumulering av lokala allelfrekvensskillnader (Jay *et al*. 2012). Detta beror på att geografiska barriärer ökar effekten av genetisk drift och grundareffekt (François *et al*. 2010). Analyser av globala data har även visat på ett riktningsberoende hos den genetiska differentieringen i mänskliga populationer; F_{ST} ökar snabbast nord-syd i Europa och Afrika och öst-väst i Asien (Jay *et al*. 2012).

Ett sätt att illustrera den genetiska diversiteten geografiskt är att lägga in cirkeldiagram över den genetiska diversiteten på en karta. Då placerar man cirkeldiagrammet på den geografiska punkt där den provtagna populationen lever, se figur 8.



Figur 8 Uppskattat kopieantal (CN) för ett antal utvalda DNA-regioner hos 52 olika populationer världen över. Omritad efter Campbell *et al*. (2011)

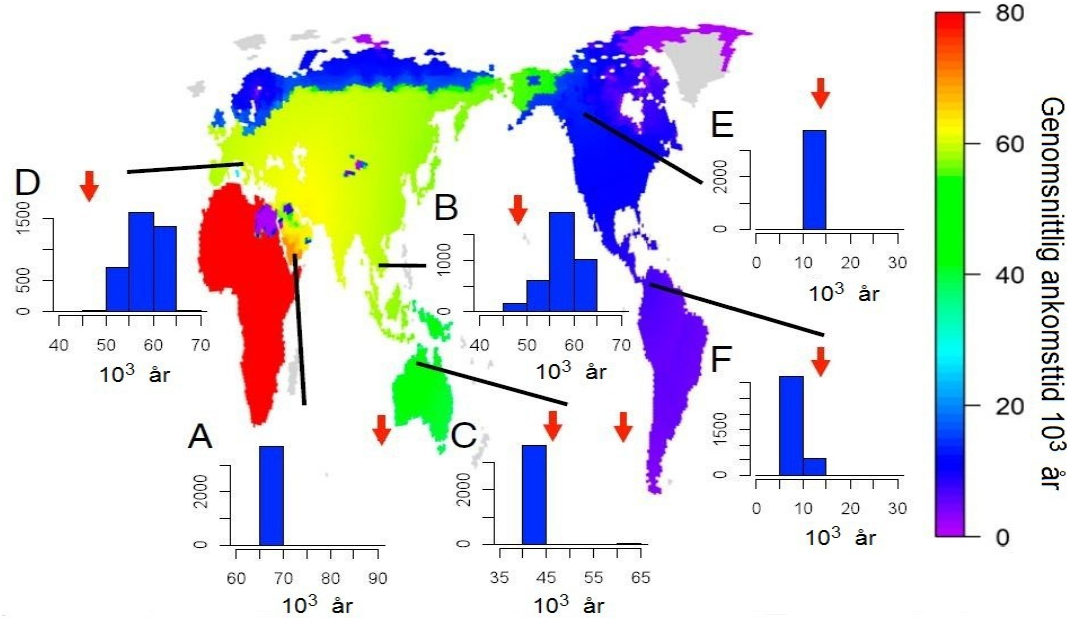
Ett annat sätt är att skriva ut de distinkta polymorfismer som identifierar varje provtagen population, se figur 9.



Figur 9 Polymorfismer på mtDNA för moderna människor (i svart), Neanderthalis (blått) och Denisova (rött). Omritad efter Krause *et al*. (2010).

Ett sätt att applicera geografisk information på genetisk variation är att använda sig av en rutnätsmodell, där varje cell har geografiska koordinater. Varje cell får även ett friktionsvärde som representerar i hur hög grad populationsstorleken uppfyller cellens bärarkapacitet. Populationen växer tills dess att den nått cellens bärarkapacitet, varvid ett antal individer migrerar för att kolonisera närmaste granncell. Modellen beror på tillväxthastigheten *r*, bärarkapaciteten *C* och migrationshastigheten *m*. Francois *et al*. (2010) inverterade *m* relaterat till de grannliggande cellernas friktionsvärde så att individer migrerar långsammare till grannceller med högt friktionsvärde och snabbt till grannceller med lågt friktionsvärde. För att skapa fylogenetiska släktträd utifrån denna modell söker man koalescenspunkter. Genom att integrera tiden *t* (antal generationer) i modellen kan man använda populationsstorleken *Ct* för en given cell för att beräkna sannolikheten för koalescens, förening, av ett par gener från den cellen. Man använder antalet migranter som kommer från närliggande celler för att beräkna sannolikheten för migration bakåt i tiden (François *et al*. 2010).

Mönster i den genetiska variationen i mänskliga populationer kan användas för att bestämma rimligheten i tidigare demografiska parametrar, exempelvis koloniseringshändelser, migrationer, populationsflaskhalsar och expansioner. Den senaste och mest omfattande spatiotemporala[22] studien kring mänsklig demografisk historia baserat på genetiska data gjordes av Eriksson *et al*. (2012). De delade in världen i lika breda celler, och gav med klimatdata ett värde för varje cells bärarkapacitet vid 62 tillfällen de senaste 120 000 åren. De valde en cell i Afrika söder om Sahara som startpunkt för populationen, och lät populationen växa tills dess att den nått cellens bärarkapacitet, då ett antal individer migrerar för att kolonisera närmaste tomma, beboeliga cell. De använde sig sedan av Wright-Fishers genetiska modell, och upprätthåller en viss migration mellan grannceller. För att skapa stokastiska genetiska släktträd för de individer från olika platser som blivit provtagna spårade de släktleden bakåt i tiden generation för generation tills koalescenspunkten för hela provet nåddes (Eriksson *et al*. 2012), se figur 10.



Figur 10 Medianen för *Homo sapiens* ankomsttid, i tusen år sedan. Histogrammen visar frekvensen av tidpunkter för nyckelområden i världen: (A) Arabiska halvön (utgångspunkten från Afrika), (B) Sydöstra Asien, (C) Australien, (D) Europa, (E) Nordamerika. Röda pilar visar dateringen för de tidigaste arkeologiska fynden av anatomiskt moderna människor i vardera område. Områden färgade i grått koloniserades aldrig, antingen på grund av extrema väderförhållanden eller brist på kopplingar till fastlandet (resor på över 100 kilometer över vatten tilläts inte i modellen). Omritad från Eriksson *et al*. (2012).

22 Rumslig och tidsmässig

Diskussion

MitokondrieDNA och Y-kromosomsDNA används för att skapa uniparental fylogeni (på mödernet respektive fadernet) och autosomalt DNA för att fastställa regionala skillnader. Skillnader vilka utgör markörer är SNP, VNTR och strukturella variationer. Den grundläggande modellen som ligger till grund för slutsatser kring mänsklig migration är en seriell grundarmodell som beskrivs antingen med hjälp av kopplingsjämvikt eller Wrights ö-modell. De huvudsakliga metoder som används för att analysera och strukturera genetiska data är fylogeni och demografi. Fylogenetiska studier görs utifrån genetisk distans eller icke-rekombinanta haplotyper. Demografiska studier bygger på mikrosatellitvarians eller heterozygotigrad.

Modellerna över evolutionära processer har blivit kritiserade för att inte representera verkligheten. Wright (1951) konstaterar att en slags labil balans mellan mutationshastighet, inavel, korsningsavel och selektion ger en mer effektiv evolutionär mekanism än enbart en av faktorerna. Han erkänner att betydande gener har pleiotropiska effekter[23] men belyser samtidigt betydelsen av genetisk kvantitativ variation.

Balaresque *et al*. (2007) menar att den genetiska diversiteten även ska undersökas utifrån perspektivet av människans evolutionära anpassning. Genom att studera selekterade gener och ta reda på fenotypiska produkter kan man svara på ytterligare frågor kring människans ursprung och spridning. Några exempel är genetiska studier på anpassningar till patogener, klimat, kost och möjligen kognition. Forskargruppen anser att studier på enskilda loci bara fångar en beskrivning av de evolutionära processerna. Ett annat problem med att sammanställa och jämföra resultat är att forskare ofta använder sig av olika material, det vill säga DNA från olika individer och populationer. När de globala mönstren för den neutrala genetiska diversiteten är kartlagda kan man fortsätta forskningen för att särskilja regionala genetiska skillnader som kan ha uppstått till följd av kulturella barriärer.

Vogel och Motulsky (1997) framhåller att populations- och evolutionär genetik ofta blir underskattade i forskningen, som fokuserar mer på medicinska problem. Eftersom mänsklig genetik ligger så nära oss finns en tendens att fokusera på praktiska tillämpningar i högre grad än grundläggande forskning. I miljön av dagens forskningsteknologi och stora forskargrupper bedrivs forskning till stor del med finansiärernas förväntningar i åtanke (Donner 2003).

Ett problem när man simulerar mänskliga migrationsmönster är svårigheterna att räkna med händelser som sker på små geografiska skalor, exempelvis fluktuationer i bärarkapacitet beroende på förändringar i resurstillgång eller miljöförändringar (François *et al*. 2010).

Datering av migrationer kan bara uppskattas av genetiska studier. Oppenheimer (2012) menar att man för att få mer exakta dateringar bör använda sig av paleoklimatologi, paleoantropologi och arkeologi. För att fullständigt kunna karaktärisera den mänskliga genetiska evolutionen krävs kunskap om hela det mänskliga genomets diversitet (Underhill & Kivisild 2007). Eriksson *et al*. (2012) menar att framtida arbete borde koncentreras på regionala studier, vilket skulle kräva detaljerade rekonstruktioner av klimat, så som nederbörd och temperatur, för att erhålla korrekta tidpunkter och platser. Då kan vi dra slutsatser kring mer specifika demografiska händelser, exempelvis teknikutveckling.

Wells (2003) konstaterar att arbetet med att utforska släktbanden mellan jordens befolkningsgrupper är ett arbete mot tiden. Den ökade globaliseringen leder till att människor migrerar i en utsträckning som aldrig skådats tidigare, och en uppluckring av nationella och kulturella gränser leder till en

23 Multipla oberoende verkningar (Nationalencyklopedin 2013e)

omfattande blandning av den mänskliga arvsmassan. I och med den ökade globaliseringen och urbaniseringen runt om i världen är de genetiska forskarna nödgade att söka data från de populationer som fortfarande lever förhållandevis isolerat. För att vi ska kunna dra slutsatser kring människans tidigare vandringar krävs att vi kan jämföra genetisk variation mellan distinkta geografiska populationer. Denna nödvändighet är på väg att försvinna till följd av något som nog de flesta av oss ändå ser som en positiv samhällsutveckling med kärlek över etniska gränser och ökad mångfald.

Datasimulering har tidigare haft en begränsad del i evolutions- och populationsgenetiken. Nu har nya lätthanterliga mjukvarupaket gjort datasimulering till ett alternativ för många forskare inom fältet (Hoban *et al*. 2012). Exempelvis använde sig Tishkoff i sin studie *The Genetic Structure and History of Africans and African Americans* (2009) sju olika mjukvarupaket (Tishkoff *et al*. 2009).

Möjligheten att extrahera och sekvensera DNA från arkeologiska lämningar ger förhoppningar om att bringa ytterligare klarhet i människogruppers släktskap och migrationer. Genom att undersöka forntida DNA från arkeologiska fynd har man kunnat jämföra mtDNA från den anatomiskt moderna människan med det från Neandertalare och Denisova homininer (Krause *et al*. 2010). Mendez *et al*. (2013) har nyligen publicerat en artikel där de ifrågasätter åldern på den moderna människans senaste gemensamma anfader utifrån Y-kromosoms DNA. Tidigare har koalescenspunkten uppskattats till runt 200 000 år sedan. Mendez forskargrupp föreslår tidpunkten till 338 000 år sedan. Deras studie av tidigare oupptäckta SNP markörer på en modern människas NRY är ett tydligt exempel på att det fortfarande finns mycket kvar att upptäcka kring människans förflutna med hjälp av de genetiska metoderna.

I arbetet med människans DNA finns flera etiska frågeställningar. Vilket ansvar har exempelvis forskare att rapportera möjliga genetiska sjukdomar hos individer i sina prov? Att inte skada, samt respekt, välgörenhet och ömsesidighet, är viktiga principer som borde ingå i en etisk handlingsplan vid sådana fynd (Segerdahl 2013).

The Human Genome projects förslag att börja samla in DNA från de mänskliga folkslag som stod på randen till utdöende skapade hätsk diskussion kring de enorma summor som skulle läggas på projektet, av vilket inget skulle gå till de deltagande individerna. Arbetet blev beryktat för sin hårdföra exploatering av utsatta grupper och omtalat som ”vampyrprojektet” (Jobling *et al*. 2004). Idag finns strikta regelverk för insamling av mänskligt DNA: deltagandet måste vara helt frivilligt och utan någon typ av ersättning, deltagaren måste vara fullt informerad om studiens syfte och garanterad om att forskningen inte kan leda till någon ekonomisk vinst för forskargruppen. Goda relationer mellan forskare och donerande populationer ligger till grund för insamlingen av DNA i dagens forskning.

Förutom att stilla nyfikenheten kring vårt ursprung kan den mänskliga genetiska forskningen leda till medicinska fördelar (genom kartläggningen av genetiska sjukdomar) och forensisk utveckling (i faderskapstest eller i sökandet efter brottslingar). Idag samlas och analyseras mänskligt DNA av flera organisationer i databaser. De har olika inriktningar. Vissa gör DNA-prov från populationer över hela världen allmänt tillgängliga, exempelvis CEPH-HGDP (Centre d’Étude du Polymorphisme Humain – Human Genome Diversity Panel 2011), medan andra erbjuder civila att bidra med sitt DNA i form beställningsbara provkit, exempelvis The Genographic Project (National Geographic Society 2013). Det finns också arbeten med uttalat medicinskt syfte, exempelvis det internationella HapMap projektet (The International HapMap 3 Consortium 2010).

Tack

Till Mattias Jakobsson som tog sig tid för att ge mig nya perspektiv och delge sina erfarenheter från fältet. Till Martin Svenda, för utomordentlig och tålmodig handledning. Till Katariina för en lärorik kurs och till mina stöttande kurskamrater.

Referenser

Ardlie KG, Kruglyak L, Seielstad M. 2002. Patterns of linkage disequilibrium in the human genome. Nature Reviews Genetics **3**: 299–309.

Campbell CD, Sampas N, Tsalenko A, Sudmant PH, Kidd JM, Malig M, Vu TH, Vives L, Tsang P, Bruhn L, Eichler EE. 2011. Population-Genetic Properties of Differentiated Human Copy-Number Polymorphisms. American Journal of Human Genetics **88**: 317–332.

Caufield PW, Saxena D, Fitch D, Li Y. 2007. Population Structure of Plasmid-Containing Strains of Streptococcus mutans, a Member of the Human Indigenous Biota. Journal of Bacteriology **189**: 1238–1243.

DeGiorgio M, Degnan JH, Rosenberg NA. 2011. Coalescence-Time Distributions in a Serial Founder Model of Human Evolutionary History. Genetics **189**: 579–593.

Donner K. 2003. De biologiska idéernas evolution. WWW-dokument 2003-11-30: http://www.svd.se/kultur/understrecket/de-biologiska-ideernas-evolution_124001.svd. Hämtad 2013-02-21.

Eriksson A, Betti L, Friend AD, Lycett SJ, Singarayer JS, Cramon-Taubadel N von, Valdes PJ, Balloux F, Manica A. 2012. Late Pleistocene climate change and the global expansion of anatomically modern humans. PNAS **109**: 16089–16094.

François O, Currat M, Ray N, Han E, Excoffier L, Novembre J. 2010. Principal Component Analysis under Population Genetic Models of Range Expansion and Admixture. Molecular Biololgy Evolution **27**: 1257–1268.

Fullerton M. 1997. The History and Geography of Human Genes. Abridged Paperback Edition. By L. Luca Cavalli-Sforza, Paolo Menozzi and Alberto Piazza. Princeton University Press, Princeton, New Jersey, 1996. Pp. 413. £25.00. ISBN 0 691 02905 9. Annals of Human Genetics **61**: 463–467.

Gillespie JH. 1998. Population Genetics - A Concise Guide. The Johns Hopkins University Press, Baltimore, London.

Griffiths AJF, Wessler SR, Lewontin RC, Carroll SB. 2008. Introduction to Genetic Analysis. 9:e uppl. W. H Freeman and Company, New York.

Jakobsson M, Scholz SW, Scheet P, Gibbs JR, VanLiere JM, Fung H-C, Szpiech ZA, Degnan JH, Wang K, Guerreiro R, Bras JM, Schymick JC, Hernandez DG, Traynor BJ, Simon-Sanchez J, Matarin M, Britton A, Leemput J van de, Rafferty I, Bucan M, Cann HM, Hardy JA, Rosenberg NA, Singleton AB. 2008. Genotype, haplotype and copy-number variation in worldwide human populations. Nature **451**: 998–1003.

Jobling M, Hurles M, Tyler-Smith C. 2004. Human Evolutionary genetics. Origins, Peoples & Disease. Garland Publishing, New York.

Kimura M, Crow JF. 1964. The Number of Alleles That Can Be Maintained in a Finite Population. Genetics **49**: 725–738.

Krause J, Fu Q, Good JM, Viola B, Shunkov MV, Derevianko AP, Pääbo S. 2010. The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. Nature **464**: 894–897.

Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, Cann HM, Barsh GS, Feldman M, Cavalli-Sforza LL, Myers RM. 2008. Worldwide Human Relationships Inferred from Genome-Wide Patterns of Variation. Science **319**: 1100–1104.

Malmquist J. 2013. Genetiska koden. WWW-dokument 2013: http://www.ne.se.ezproxy.its.uu.se/lang/genetiska-koden. Hämtad 2013-03-17.
Mendez FL, Krahn T, Schrack B, Krahn A-M, Veeramah KR, Woerner AE, Fomine FLM, Bradman N, Thomas MG, Karafet TM, Hammer MF. 2013. An African American Paternal Lineage Adds an Extremely Ancient Root to the Human Y Chromosome Phylogenetic Tree. The American Journal of Human Genetics **92**: 454–459.
National Geographic Society. 2013. About the Genographic Project - National Geographic. WWW-Dokument 2013: https://genographic.nationalgeographic.com/about/. Hämtad 2013-02-04.
Nationalencyklopedin. 2013g. RFLP. WWW-dokument 2013: http://www.ne.se.ezproxy.its.uu.se/rflp. Hämtad 2013-03-05.
Nationalencyklopedin. 2013c. Heterozygoti. WWW-dokument 2013: http://www.ne.se.ezproxy.its.uu.se/lang/heterozygoti?i_h_word=heterosis. Hämtad 2013-03-04.
Nationalencyklopedin. 2013e. Pleiotropi. WWW-dokument 2013: http://www.ne.se/pleiotropi. Hämtad 2013-03-04.
Ogurtsov AY, Sunyaev S, Kondrashov AS. 2004. Indel-Based Evolutionary Distance and Mouse–Human Divergence. Genome Research **14**: 1610–1616.
Ohta T, Kimura M. 1973. A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. Genetics Research **22**: 201–204.
Oppenheimer S. 2012. Out-of-Africa, the peopling of continents and islands: tracing uniparental gene trees across the map. Philosophical transactions of the royal society Biology **367**: 770–786.
Ovesen J. 2013. Raser Historik. Nationalencyklopedin. WWW-dokument 2013: http://www.ne.se.ezproxy.its.uu.se/lang/m%C3%A4nniskan?i_h_word=moderna %20syntesen. Hämtad: 2013-03-07.
Pettersson U. 2013. Avläsning av DNA-molekylens sammansättning (DNA-sekvensering). Nationalencyklopedin. WWW-dokument 2013: http://www.ne.se.ezproxy.its.uu.se/genteknik/avl%C3%A4sning-av-dna-molekylens-sammans%C3%A4ttning-dna-sekvensering. Hämtad 2013-03-07.
Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW. 2002. Genetic Structure of Human Populations. Science **298**: 2381–2385.
Segerdahl P. 2013. Ethical principles causing moral hallucinations. WWW-Dokument 2013-02-20: http://ethicsblog.crb.uu.se/. Hämtad 2013-02-23.
The International HapMap 3 Consortium. 2010. Integrating common and rare genetic variation in diverse human populations. Nature **467**: 52–58.
Tishkoff SA, Reed FA, Friedlaender FR, Ehret C, Ranciaro A, Froment A, Hirbo JB, Awomoyi AA, Bodo J-M, Doumbo O, Ibrahim M, Juma AT, Kotze MJ, Lema G, Moore JH, Mortensen H, Nyambo TB, Omar SA, Powell K, Pretorius GS, Smith MW, Thera MA, Wambebe C, Weber JL, Williams SM. 2009. The Genetic Structure and History of Africans and African Americans. Science **324**: 1035–1044.
Underhill PA, Kivisild T. 2007. Use of Y Chromosome and Mitochondrial DNA Population Structure in Tracing Human Migrations. Annual Review Genetics **41**: 539–564.
Wells S. 2003. The Journey of Man. A Genetic Odyssey. Penguin Group, London, England.ns ofHuman Genetic Differentiation. Molecular Biology Evolution, doi 10.1093/molbev/mss259.
Jobling M, Hurles M, Tyler-Smith C. 2004. Human Evolutionary genetics. Origins, Peoples & Disease. Garland Publishing, New York.
Kimura, M., 1970. The length of time required for a selectively neutral mutant to reach fixation through random frequency drift in a finite population. Genetics Research 15, 131–133.
Kimura, M., Crow, J.F., 1964. The Number of Alleles That Can Be Maintained in a Finite

Population. Genetics 49, 725–738.
Kimura M, Weiss GH. 1964. The Stepping Stone Model of Population Structure and the Decrease of Genetic Correlation with Distance. Genetics **49**: 561–576.
National Geographic Society. 2013. About the Genographic Project - National Geographic. WWW-dokument 2013: https://genographic.nationalgeographic.com/about/. Hämtad 2013-02-04.
Nationalencyklopedin. 2013a. Heterozygoti. WWW-dokument 2013: http://www.ne.se.ezproxy.its.uu.se/lang/heterozygoti?i_h_word=heterosis. Hämtad 2013-03-04.
Nationalencyklopedin. 2013b. Metapopulation. WWW-dokument 2013: http://www.ne.se/lang/metapopulation. Hämtad 2013-03-04.
Nationalencyklopedin. Pleiotropi. 2013c. WWW-dokument 2013: http://www.ne.se/pleiotropi. Hämtad 2013-03-04.
Ogurtsov AY, Sunyaev S, Kondrashov AS. 2004. Indel-Based Evolutionary Distance and Mouse–Human Divergence. Genome Research **14**: 1610–1616.
Ohta, T., Kimura, M., 1973. A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. Genetics Research 22, 201–204.
Oppenheimer S. 2012. Out-of-Africa, the peopling of continents and islands: tracing uniparental gene trees across the map. Philosophical transactions of the royal society Biological Sciences **367**: 770–786.
The International HapMap 3 Consortium. 2010. Integrating common and rare genetic variation in diverse human populations. Nature **467**: 52–58.
Tishkoff SA, Reed FA, Friedlaender FR, Ehret C, Ranciaro A, Froment A, Hirbo JB, Awomoyi AA, Bodo J-M, Doumbo O, Ibrahim M, Juma AT, Kotze MJ, Lema G, Moore JH, Mortensen H, Nyambo TB, Omar SA, Powell K, Pretorius GS, Smith MW, Thera MA, Wambebe C, Weber JL, Williams SM. 2009. The Genetic Structure and History of Africans and African Americans. Science **324**: 1035–1044.
Underhill PA, Kivisild T. 2007. Use of Y Chromosome and Mitochondrial DNA Population Structure in Tracing Human Migrations. Annual Review Genetics **41**: 539–564.
Wells S. 2003. The Journey of Man. A Genetic Odyssey. Penguin Group, London.
Wright S. 1951. Fisher and Ford on “The Sewall Wright Effect”. American Scientist **39**: 452–479.