

Molekyyli populaatiogenetiikka

Hedrick 2005, kappale 8, pp. 452-462, 428-449

- Demografian ja valinnan vaikutukset koalesenssipuihin
- Valinnan havaitseminen
 - Tajiman D
 - Divergenssin ja polymorfismin vertailu
 - Valinnan pyyhkäisy ja diversiteetti

Original slides by Outi Savolainen

Koalesenssipuut ja demografia, Hedrick 2005

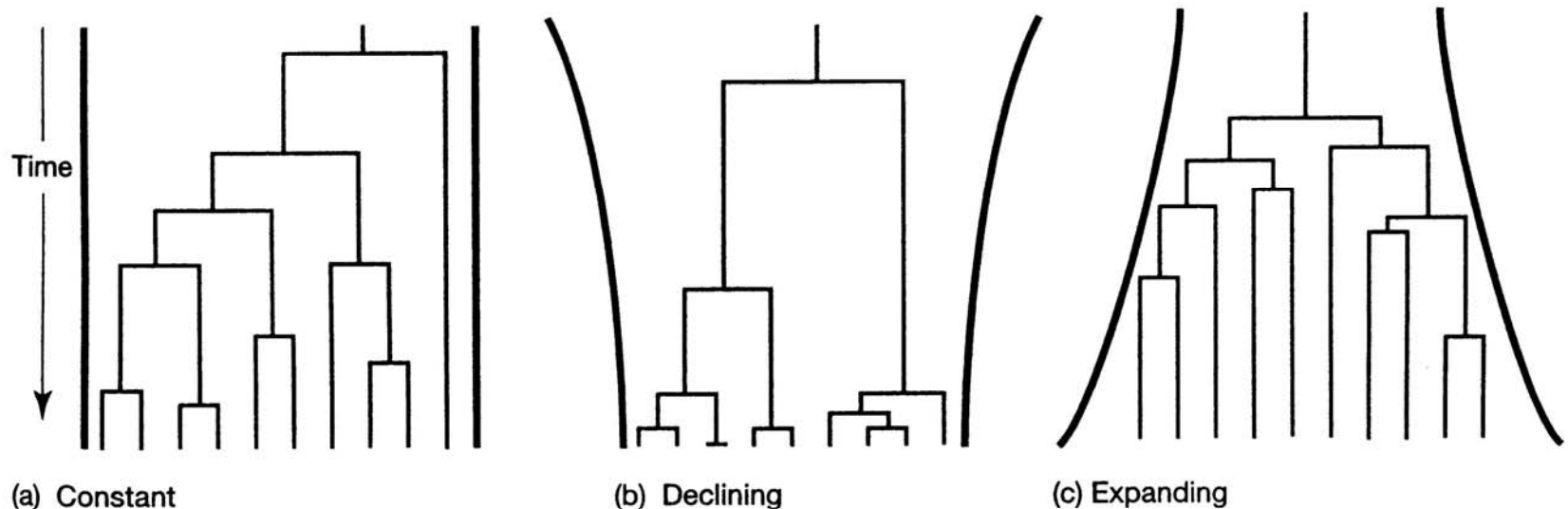


Figure 8.18. The theoretical distributions of coalescent times in genealogies with 10 contemporary samples under three scenarios of historical population change: (a) constant population size, (b) declining population size, and (c) increasing population size (Garrigan *et al.*, 2002). The distance between the thick lines indicates the relative population sizes at different times, and the genealogies reflect coalescent events in periods of

Koalesenssin laajentaminen

- Pariutumissysteemit
- Populaatiokoon muutokset
 - eksponentiaalinen, logistinen, satunnainen, pullonkaulat
- Populaatorakenne
 - saarimalli, maantieteelliseen rakenteeseen perustuvat mallit, jatkuvat mallit
- Valinta
 - tasapainottava – kaksi- ja monialleelinen
 - suuntaava – adaptiivinen ja haitallinen
- Rekombinaatio

Tärkeitä viitteitä

- Hein, J. Schierup, M. Wiuf, C. 2004 Gene genealogies, variation and evolution. Oxford Univ. Press.
- Stumpf, M and McVean, G. A. T. 2003. Estimating recombination from population genetic data. Nature Reviews Genetics 4: 959-968.
- Rosenberg, N. and Nordborg, M. 2002. Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. Nature Reviews Genetics 3: 380-390.
- Kingman, J. F. C. 1982. The coalescent. Stochastic Proc. Appl. 13: 1461-1463.
- Hudson, R. R. 1983. Testing the constant-rate neutral allele model with protein sequence data. Evolution 37: 203-217.
- Hartl, D. H. and Clark, A. G. 2007. Principles of Population Genetics. 4th edition. Sinauer.
- www.coalescent.dk

	899 ins G	904 T	908 A	908/09 del —	919 G	1098 T	1114 G	1279 Phe/Ile T	1303 Val/Leu G	1377/93 del TAGAATTCTAATTT
Landsberg	*	A	C	*	A	*	A	*	*	*
Al-0	*	A	C	*	A	*	A	*	*	*
Bs-1	*	A	C	*	A	*	A	*	*	*
Mt-0	*	A	C	*	A	*	A	*	*	*
Yo-0	*	A	C	*	A	*	A	*	*	*
Shokei	*	A	C	*	A	G	A	*	*	*
Bl-1	*	*	*	*	*	*	*	*	*	*
Gr-1	*	*	*	*	*	*	*	*	*	*
Col-0	*	A	C	*	A	*	*	*	*	*
Es-0	*	*	*	*	*	*	*	*	*	*
Chi-0	*	*	*	*	*	*	*	*	*	*
Pog-0	*	*	*	*	*	*	*	*	*	*
Aa-0	—	*	*	T	*	*	*	*	*	*
Ita-0	*	*	*	*	*	*	*	*	C	—
Ci-0	*	*	*	*	*	*	*	*	C	—
Bla-10	*	*	*	*	*	*	*	*	C	—
Hiroshima	*	*	C	*	*	*	A	A	C	—

Innan et al. 1996 Genetics143:1761. Adh in ecotypes of Arabidopsis thaliana nt 899-1398 ~500 nt

Sites	112	139	243	427	628	762	915	1156	1456	1687	2029	2038
Consensu	C	T	C	C	T	T	A	C	G	C	T	T
Bro1	.	.	.	.	.	.	.	.	.	.	.	G
Bro2	.	.	.	.	.	.	.	.	A	.	.	.
Bro3	.	.	.	.	C	.	.	.	.	T	.	G
Bro4	.	.	.	T	.	.	.	.	.	.	.	.
Bro5	.	.	.	.	.	.	.	.	.	T	G	.
Kol1	.	.	.	.	C	.	.	G	.	.	.	G
Kol2	.	.	.	.	.	.	.	.	.	T	.	.
Kol3	.	.	T	.	.	.	.	.	.	T	.	.
Kol4	.	.	T	.	.	.	.	G	.	.	.	.
Kol5	.	.	.	.	C	.	.	.	.	.	.	.
Kir19	.	C	.	.	.	G	.	.	.	.	.	.
Kir21	.	.	.	T	.	.	.	.	.	.	.	.
Kir23	T	.	.	.	.	.	.	.	.	.	.	.
Kir27	.	C	.	.	.	.	.	G	.	.	.	.
Kir30	T	.	.	.	.	.	T	.	.	T	.	G
Lillo1	.	.	.	.	C	.	.	.	.	T	.	G
Lillo3	.	.	.	.	C	.	.	.	.	T	.	G
Lillo17	.	.	.	.	C	.	.	.	.	.	.	.
Lillo19	.	.	.	.	C	.	.	.	.	.	.	.
Lillo22	.	.	.	.	C	.	.	.	.	.	.	.

Neutraalista tasapaino-
mallista poikkeavan
muuntelun havaitseminen

Adh –muuntelu, *A.*
thaliana (Innan *et*
al. 1996) 500 nt, π
=0.019

Nukleotidi-
polymorfismi *pal1*-
geenin 2043
nukleotidipaikassa
(Dvornyk *et al.*
2002)

Neutraalista tasapainomallista poikkeavan muuntelun havaitseminen

- Tajiman D
- HKA-testi
- McDonald Kreitman
- Joillain menetelmillä poikkeama neutraalimallista voi johtua joko demografiasta tai valinnasta

Tajiman testi

- Verrataan segregoituvien nukleotidipaikkojen (θ_w) tai nukleotididiversiteetin (θ_T) perusteella arvioituja θ :n estimaatteja
 - θ_w segregoituvien nukleotidipaikkojen lukumäärä, myös harvinaiset lasketaan
 - θ_T parittainen sekvenssien vertailu, keskimääräisellä frekvenssillä esiintyvä muuntelu vaikuttaa eniten
- Ero estimaattien välillä johtuu poikkeamasta neutraalista tasapainomallista
- Demografia vai valinta?

Tajiman D

$$D = \frac{\Theta_T - \Theta_W}{\text{Std}(\Theta_T - \Theta_W)}$$

Suuri arvo kertoo terminaalihaarojen olevan lyhyitä

Pieni arvo kertoo syvien haarojen olevan lyhyitä

➔ poikkeamat koalesenssipuun rakenteessa havaitaan Tajiman D -testin avulla

Nyrkkisääntö: $D > 2$ tai $D < -2$ viittaa merkitsevään poikkeamaan

Tajiman D ja HLA, Garrigan ja Hedrick 2003

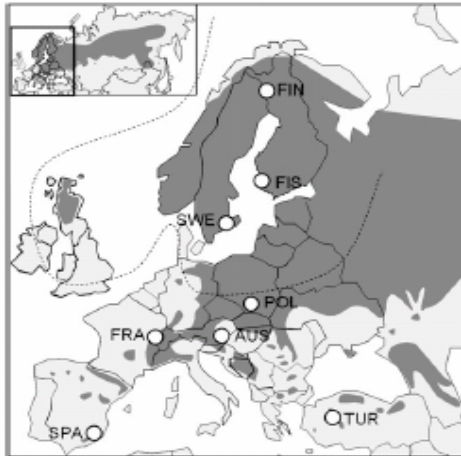
TABLE 8. Homozygosity (F) and Tajima's D statistics calculated for a global sample of *HLA-A* sequences. A total of 537 sites were examined. Simulations of the neutral genealogical process, conditioned on the observed number of segregating sites, were used to assess significance of Tajima's D statistic (significant values shown in bold). Data were obtained from the listed reference.

Population	2N	Alleles	F	Prob(F)	Tajima D	Prob(D)	Reference
Ainu	100	9	0.182	0.06	3.143	<0.01	Bannai et al. 2000
Australian	367	7	0.327	0.24	3.736	<0.01	Gao et al. 2000
Chinese	298	22	0.140	0.48	2.356	0.01	Middleton et al. 2000
French	248	20	0.162	0.60	2.440	<0.01	Bugawan et al. 2000
Havasupai	244	3	0.401	0.03	4.049	<0.01	Markow et al. 1993
Kagui	100	4	0.356	0.09	3.254	<0.01	Middleton et al. 2000
Molucca	52	7	0.214	0.09	2.514	<0.01	Bugawan et al. 1999
Omani	236	27	0.100	0.42	2.498	<0.01	Middleton et al. 2000
PNG-Lowland	94	6	0.336	0.32	2.447	<0.01	Bugawan et al. 1999
PNG-Highland	188	5	0.634	0.75	1.013	0.13	Bugawan et al. 1999
Zapotec	137	6	0.235	0.02	1.976	0.01	Hollenbach et al. 2001
Zulu	200	23	0.069	<0.01	2.441	<0.01	Middleton et al. 2000
Global	2164	45	0.126	0.73	3.691	<0.01	

Huomaa positiiviset Tajiman D :n arvot – liian vähän harvinaisia alleeleja, valinta ylläpitää neutraalitulannetta tasaisempia alleelifrekvenssejä populaatioissa.

Männyn demografia

(Pyhäjärvi *et al.* 2007 Genetics 177: 1713–1724)

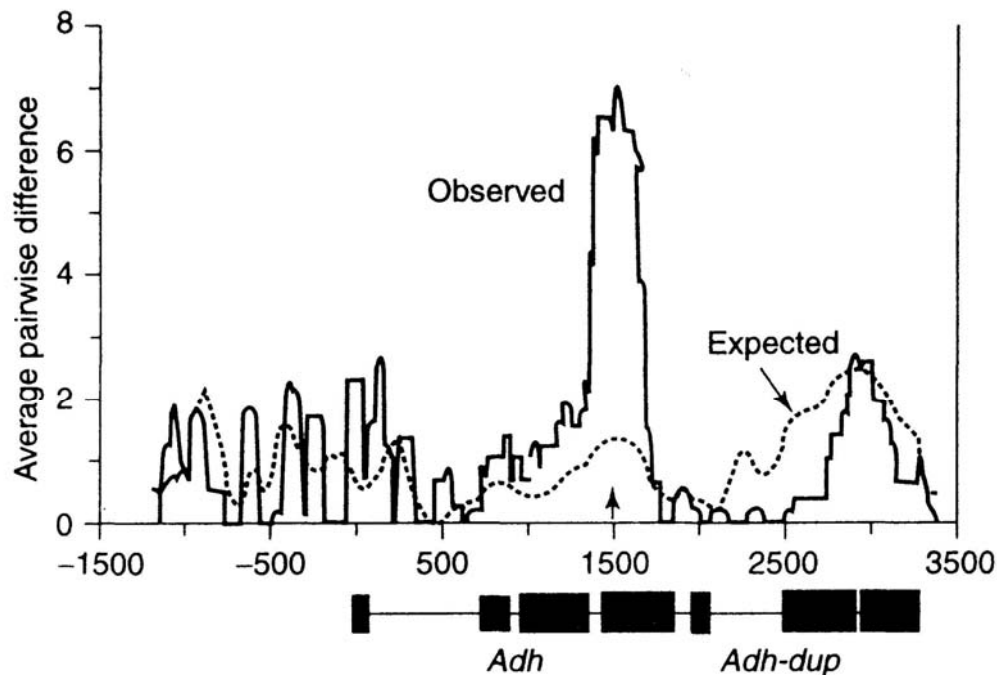


Tajiman D negatiivinen keskeisissä ja pohjoisissa populaatioissa – vanha pullonkaula ja populaatioekspansio

		Tajima's D^a	
	θ^b	Total	Silent
North	0.0056	-0.63 (1.06) **	-0.53*
Central	0.0050	-0.32 (0.99)	-0.51*
Spain	0.0047	-0.16 (1.09)	0.12
Turkey	0.0057	-0.24 (0.88)	0.21

HKA-testi, *Drosophila*:n Adh-geeni

Verrataan lajien välistä divergenssiä polymorfismiin



Neutraalisuusodotus:
divergenssi ja poly-
morfismi korreloituneita,
huomaa poikkeama
eksonissa 4

Divergenssi μ
Polymorfismi $4N\mu$

Figure 8.8. A sliding window of the observed and expected genetic variation over the 5' flanking region, the *Adh* gene, and the *Adh-dup* gene in *Drosophila* (after Hickey and Hudson, 1991). The expected values are based on a no selection

McDonald-Kreitman testi

- Verrataan divergenssiä ja polymorfismia synonyymisissä ja ei-synonyymisissä nukleotidipaikoissa

	Div	Polym
Syn		
Nonsyn		

McDonald-Kreitman test

Hedrick 2005

TABLE 8.9 The number of nonsynonymous (replacement) and synonymous substitutions for fixed differences between species and polymorphism within species (a) in general, (b) for *Adh* in three *Drosophila* species (McDonald and Kreitman, 1991), and (c) for *G6pd* in *D. melanogaster* and *D. simulans* (Eanes *et al.*, 1993). Below, data from humans and chimpanzees are given for (d) mtDNA gene *ND3* (Nachman *et al.*, 1996), (e) *G6pd* (Verrelli *et al.*, 2002), and (f) *HLA-B* (Garrigan and Hedrick, 2003).

	(a) General		(b) <i>Adh</i>		(c) <i>G6pd</i>	
	<i>Fixed</i>	<i>Polymorphic</i>	<i>Fixed</i>	<i>Polymorphic</i>	<i>Fixed</i>	<i>Polymorphic</i>
Nonsynonymous	N_F	N_P	7	2	21	2
Synonymous	S_F	S_P	17	42	26	36
Ratio	$N_F/S_F = N_P/S_P$		0.41	0.05	0.81	0.06
	(c) <i>ND3</i>		(d) <i>G6pd</i>		(e) <i>HLA-B</i>	
Nonsynonymous	4	8	0	5	0	76
Synonymous	31	10	44	23	0	49
Ratio	0.13	0.8	0.0	0.28	—	1.61

McDonald and Kreitman (1991) applied this test to data from coding region of *Adh* from *Drosophila* species *D. melanogaster*, *D. simulans*, and *D. yakuba*. Table 8.10 gives the nucleotide sequence for the nonsynonymous (replacement) differences for the three species and the status as far as fixed differences between species (seven sites) and polymorphic variation within species (two sites). For example, position 781 is considered a fixed difference

Valinnan havaitseminen

1710

D. GARRIGAN AND P. W. HEDRICK

TABLE 1. The tests of selection and their evolutionary parameters for detecting selection on the three different timescales.

Timescale	Test	Parameters
Current generation	Hardy-Weinberg	Selection coefficient, number of alleles
	Mendelian proportions	Selection coefficient
	Random mating	Selection coefficient
	Fitness associations	Selection coefficient
Recent past	Ewens-Watterson	Population size, number of alleles, selection coefficient and duration
	Linkage disequilibrium	Population size, recombination rate, selection coefficient and duration
	Geographical congruency	Population size, population structure, selection coefficient and duration
Distant past	Nonsynonymous/synonymous substitution rate	Selection coefficient and duration, mutation rate
	McDonald-Kreitman	Selection coefficient and duration, mutation rate, species divergence times
	Tajima's D	Population size, population structure, mutation rate, selection coefficient and duration
	Transspecies polymorphism	Selection coefficient, mutation rate, species divergence times

Valinnan pyyhkäisy, hitchhiking

- Suuntaava valinta vähentää muuntelua
- Esimerkki: malarialoisen *Plasmodium falciparum* kiniiniresistenssilokus
 - lääkettä alettiin käyttää n. 40 vuotta sitten
 - resistenssiominaisuus *pfcrt* –geenissä, useita mutaatioita
 - resistenssilokuksen lähellä vähemmän muuntelua kuin muualla

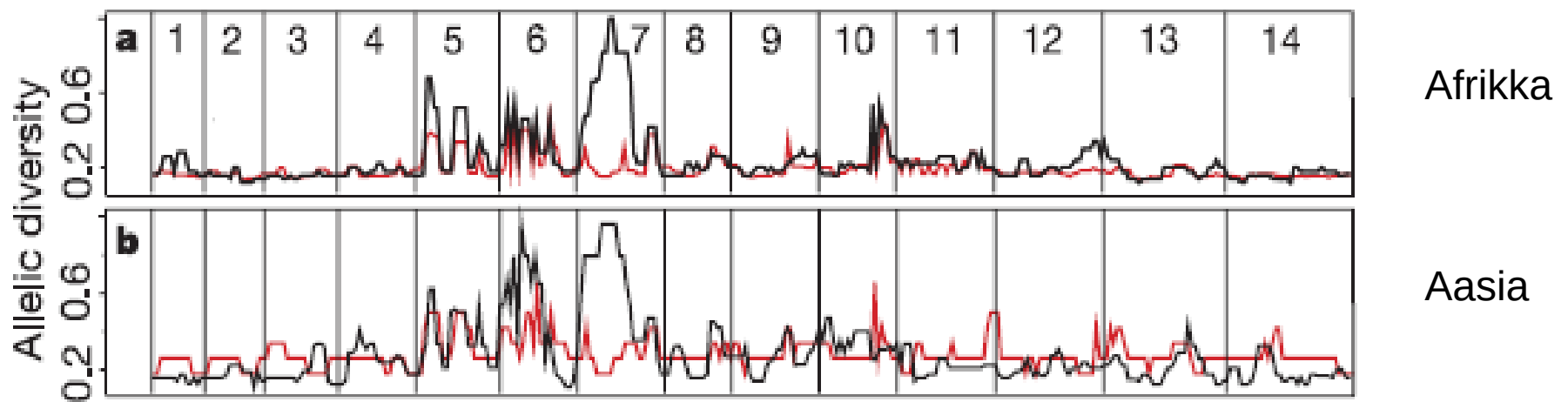


Figure 3 Genome-wide scans for loci of reduced diversity and association of reduced diversity with the CQR phenotype. The numbered panels represent the 14 chromosomes covered by 342 markers. **a**, **b**, Allelic diversity, plotted as $(1 - \text{median adjusted allelic diversity of five contiguous markers})$, for CQS (red) and CQR (black) isolates from Africa (**a**) and Asia (**b**, excluding CQR PNG isolates). Peaks represent regions with reduced diversity.

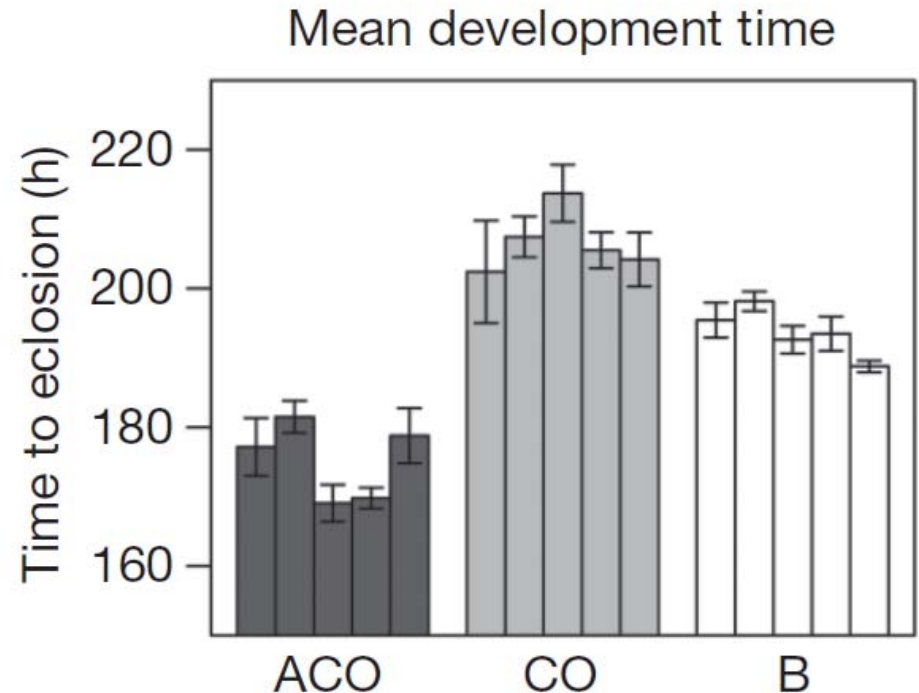
PUNAINEN – sensitiivinen, MUSTA – resistentti kanta

Huomaa: Alleelidiversiteetti vähentynyt eritoten kromosomissa 7, missä resistenssilokus sijaitsee, tärkeää verrata genomiseen tasoon (esim. demografian vaikutus)

Wootton *et al.* 2002 Nature 418, 320-323.

Kokeellinen evoluutio *Drosophila*:lla – pehmeä valinnan pyyhkäisy, Burke *et al.* 2010 Nature 467: 587–590.

- *D. melanogaster* –linjoja valittiin nopean yksilönkehityksen suhteen 600 sukupolvea
- tutkittiin geneettisiä muutoksia
- uusien mutaatioiden täydellinen fiksoituminen eri linjoissa TAI
- olemassa olevaan muunteluun perustuva valinta



ACO – valinta, nopea kehitys
CO - kontrolli
B - peruspopulaatio

Tulokset, Burke *et al.* 2010

Täydellinen valinnan pyyhkäisy uuden hyödyllisen mutaation tapauksessa: kussakin replikaattilinjassa oma mutaatio joka fiksoituu – monomorfisia alueita, jotka tn. eroavat linjojen välillä

Vaste perustuu olemassa olevaan muunteluun – samat geenialueet eri linjoissa, mahdollisesti eivät täysin monomorfisia

Pehmeä valinnan pyyhkäisy, Burke *et al.* 2010

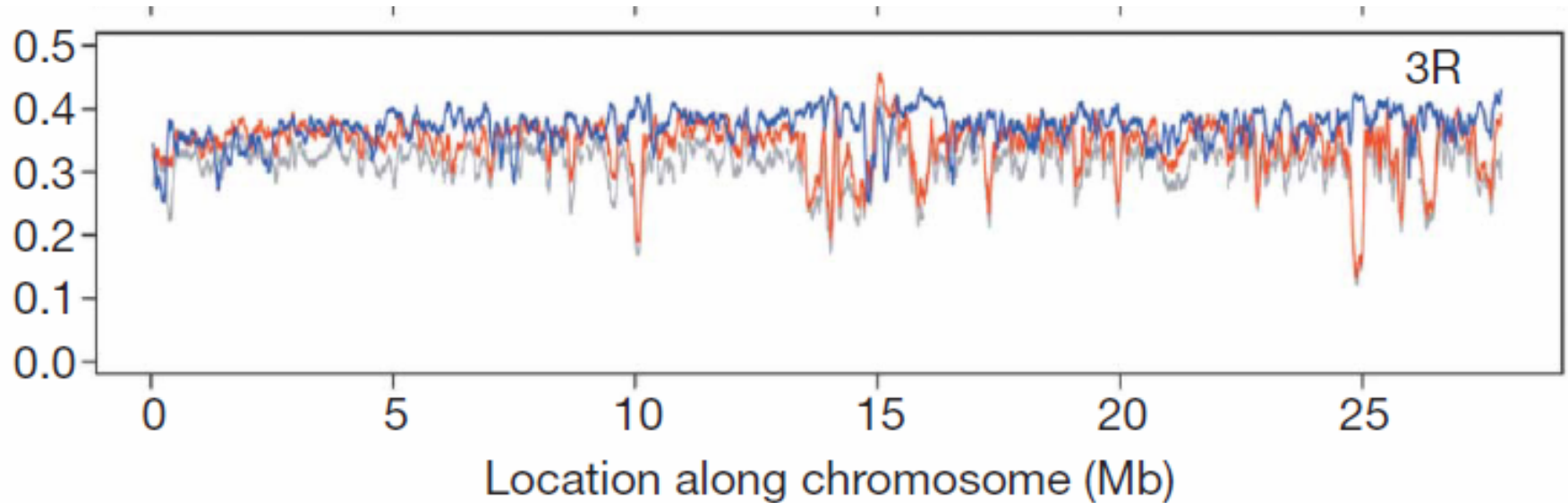


Figure 3 | Heterozygosity throughout the genome. Sliding-window analysis (100 kb) of heterozygosity in the CO pool (blue), the ACO pool (red) and ACO₁ (grey), with a 2-kb step size. The panels show the five major chromosome arms of *D. melanogaster*.

- Paikallisia heterotsygotian alenemisia
- Alhaisen heterotsygotian alueet samoja, joissa todettiin erilaiset alleelifrekvenssit valittujen ja kontrollilinjojen välillä
- Pehmeä valinnan pyyhkäisy: heterotsygotiat eivät nollassa (kuten klassisessa valinnan pyyhkäisyssä odotettaisiin)

Valinta kehitysajan suhteen, Burke *et al.* 2010

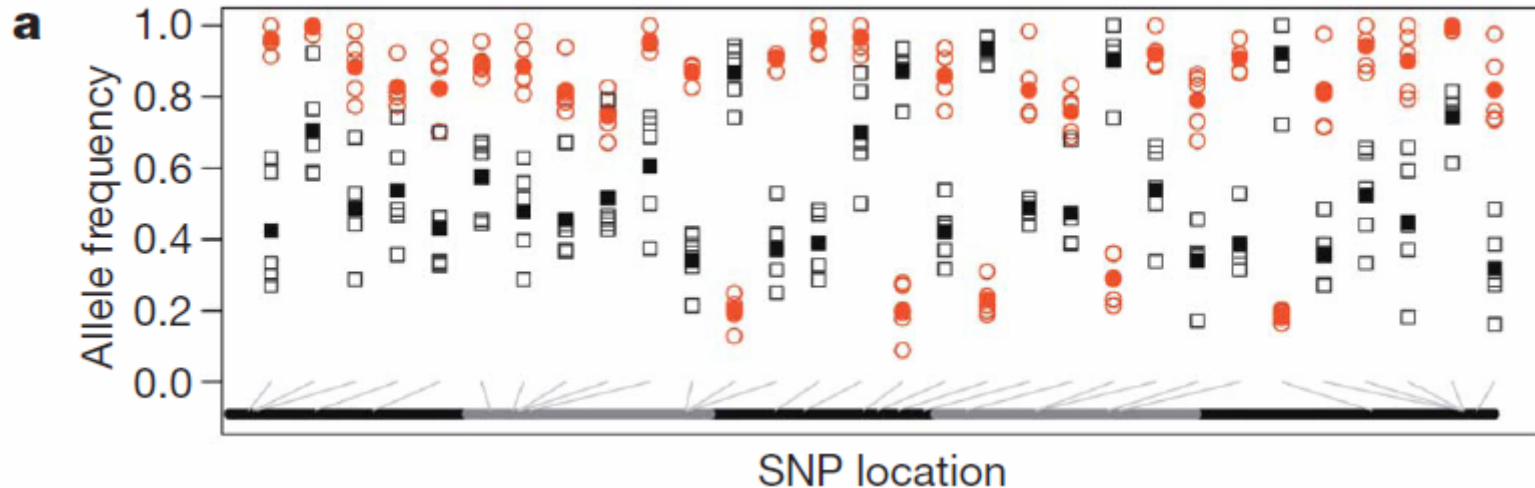


Figure 4 | Analysis of individual genotypes, measured by cleaved amplified polymorphic sequence (CAPS) techniques. a, Allele frequency estimates of the most common allele at 30 SNPs genotyped in 35 females per replicate population. Red circles represent ACO estimates and grey squares represent CO estimates. Open symbols are allele frequencies for ACO₁–ACO₅ and CO₁–CO₅, and filled symbols represent treatment means. Alternating black and grey bars designate the X, 2L, 2R, 3L, and 3R arms, respectively, with grey lines

- Alleelifrekvenssien ja heterotsygotiatasojen yhtymäkohdat toistojen välillä
- Valinta vaikuttaa samoihin keskimääräisellä frekvenssillä esiintyviin variantteihin kussakin populaatiossa