BIO334 Practical Bioinformatics
Introduction to genome-wide polymorphisms

Take BIO373 for biological implications,
or refer to the textbooks

1. Futuyma "*Evolution*", Sinauer, 2005, 2017, Chap 12
2. Hartl and Clark "*Principles of Population Genetics*", Sinauer 4th ed., 2007
3. Freeman and Herron "*Evolutionary Analysis*" 5th Ed., Pearson, 2013, Chap. 7, 8

Department of Evolutionary Biology and Environmental Studies
University of Zurich
Yasuhiro (Yasu) Sato, a senior postdoc of Shimizu Group
yasuhiro.sato@uzh.ch

Overview
molecular population genetics

Introduction:
Genome-wide polymorphism and next-generation sequencers

1. nucleotide diversity
selective sweep / hitchhiking, and balancing selection

2. Tajima's *D*

3. positive selection on replacement mutations

Human 1000 genomes were published on 28 Oct 2010: How are they meaningful?

Example: evolution of language

Was there positive selection to have language?
How many genes were necessary to obtain language?
How and when did the language evolve?

Did Neanderthal have language,
which diverged 3-400,000 years ago?

Limited evidence: archaeology, fossil of skeletons

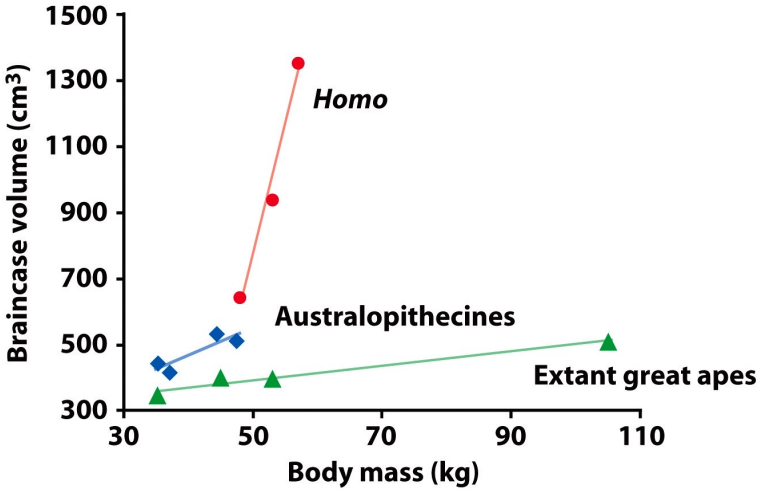


Figure 20-32 Evolutionary Analysis, 4/e
© 2007 Pearson Prentice Hall, Inc.

Figure 20-31 Evolutionary Analysis, 4/e
© 2007 Pearson Prentice Hall, Inc.

From Freeman and Herron (2007) "*Evolutionary Analysis (4th edn)*"

How to analyze multiple sequences?

```
SCR-1476  TGAGACATAA ACATTATATT TTTCAATTAG TCATTAAAAT AAAAATTTCT
SCR-1640  TGAGACATAA ACATTATATT TTTCAATTAG TCATTAAAAT AAAAATTTCT
SCR-20    TGAGACATAA ACATTATATT TTTCAATTAG TCATTAAAAT AAAAATTTCT
SCR-6090  TGAGACATAA ACATTATATT TTTCAATTAG TCATTAAAAT AAAAATTTCT
SCR-6622  TGAGACATAA ACATTATATT TTTCAATTAG TCATTAAAAT AAAAATTTCT
SCR-6665  TGAGACATAA ACATTATATT TTTCAATTAG TCATTAAAAT AAAAATTTCT
SCR-6680  TGAGACATAA ACATTATATT TTTCAATTAG TCATTAAAAT AAAAATTTCT
SCR-6780  TGAGACATAA ACATTATATT TTTCAATTAG TCATTAAAAT AAAAATTTCT
SCR-6807  TGAGACATAA ACATTATATA TTTCAATTAG TCATTAAAAT AAAAATTTCT
SCR-6876  TGAGACATAA ACATTATATT TTTCAATTAG TCATTAAAAT AAAAATTTCT
SCR-6918  TGAGACATAA ACATTATATT TTTCAATTAG TCATTAAAAT AAAAATTTCT
SCR-915   TGAGACATAA ACATTATATT TTTCAATTAG TCATTAAAAT AAAAATTTCT
SCR-917   TGAGACATAA ACATTATATT TTTCAATTAG TCATTAAAAT AAAAATTTCT
```

Contents:
molecular population genetics

0. Genome-wide polymorphism and next-generation sequencers

1. nucleotide diversity
selective sweep / hitchhiking, and balancing selection

2. Tajima's *D*

3. positive selection on replacement mutations

Nucleotide diversity π

(Nukleotid Diversität)

(θ_π, θ_T)

Basic index of molecular population genetics

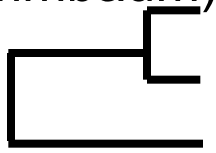
Average proportion of pairwise differences between the sequences

(Durchschnittliches Verhältnis von paarweisen Unterschieden zwischen den Sequenzen)

```
                    .*....*.*.
(1) AGGCTGCATC      AGGCTGCATC
(2) .A.........  =  AAGCTGCATC
(3) ......T.C.      AGGCTGTACC
```

```
     (1)   (2)   (3)
(1)  ---
(2)  ---   ---
(3)  ---   ---   ---
```

Genealogy
(Stammbaum)



$$\pi = \sum_{i<j} \pi_{ij} / n_c = ?$$

$$n_c = \frac{n(n-1)}{2}$$

number of pairwise comparisons
(Anzahl von paarweisen Vergleichen)

π_{ij} proportion of differences between i-th and j-th
(Verhältnis der Unterschiede zwischen i und j)

$$\pi = \sum_{i<j} \pi_{ij} / n_c = \frac{\pi_{12} + \pi_{13} + \pi_{23}}{3} = ?$$

nucleotide diversity π

Your own calculation would hopefully help intuitive understanding
(on average, 2 out of 10 are different)

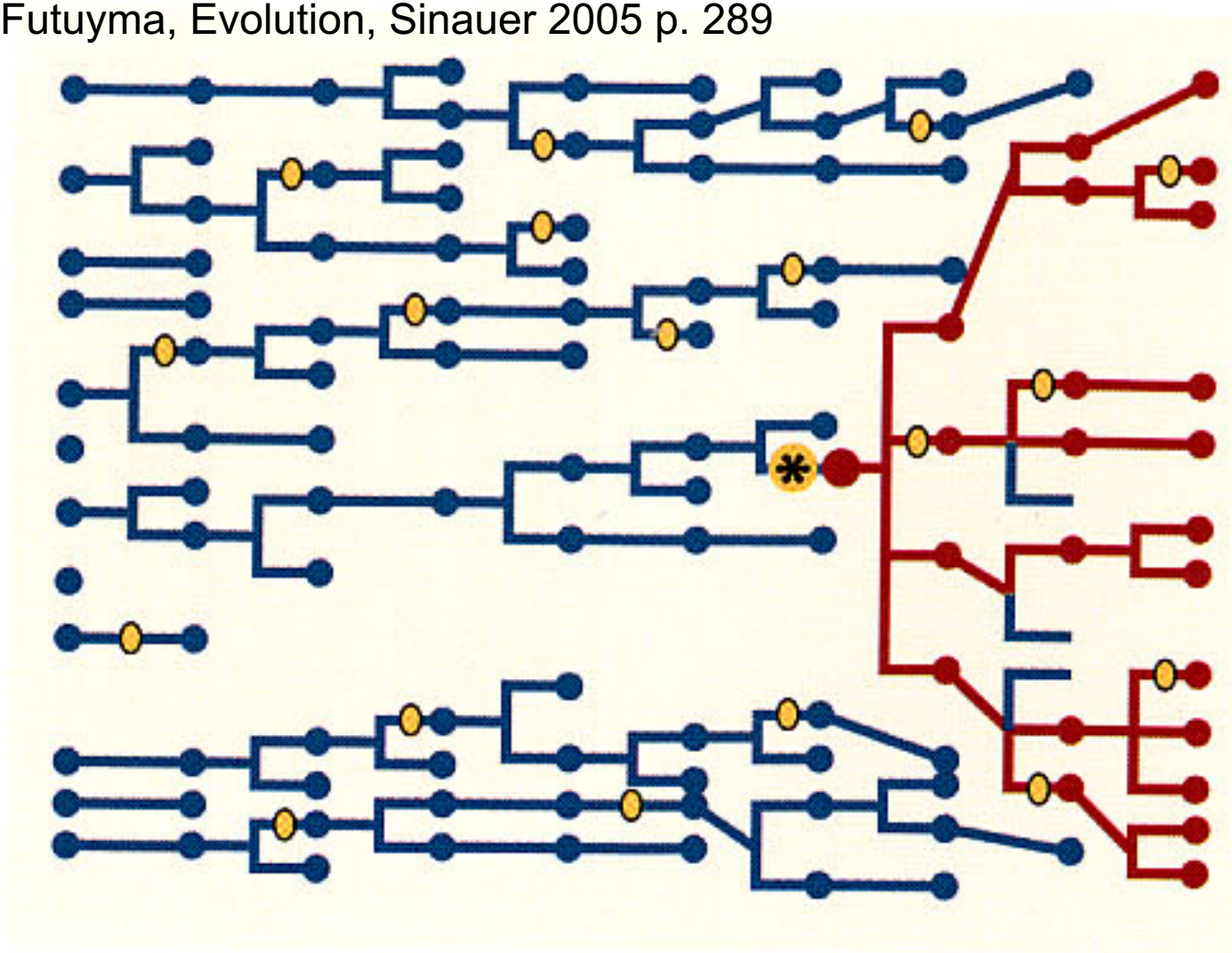
Genomic average of nucleotide diversity π

Homo sapiens: ~0.0008
Arabidopsis thaliana and *Drosophila melanogaster*: ~0.007

Coalescent theory: coalescence and mutations

(Koaleszenz bzw. Vereinigungs- Theorie – Koaleszenz und Mutationen

Futuyma, Evolution, Sinauer 2005 p. 289



AGGCTGCATC
.....T...
.....

.A.T.....
.A.....
.....
.....
.....C....
.....
.....
..T.....
..T.....

Natural selection can be classified into 2 types based on the pattern of polymorphism

1. positive selection

(directional selection, Darwinian selection):
selection for an allele that increases fitness.
Positive selection results in selective sweep.

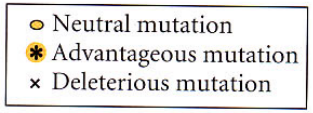
2. balancing selection

a form of natural selection that maintains polymorphism at a locus within a population.
This includes heterozygosity advantage, negative frequency dependent selection, local adaptation, etc. Diversifying selection can be used in a similar meaning.

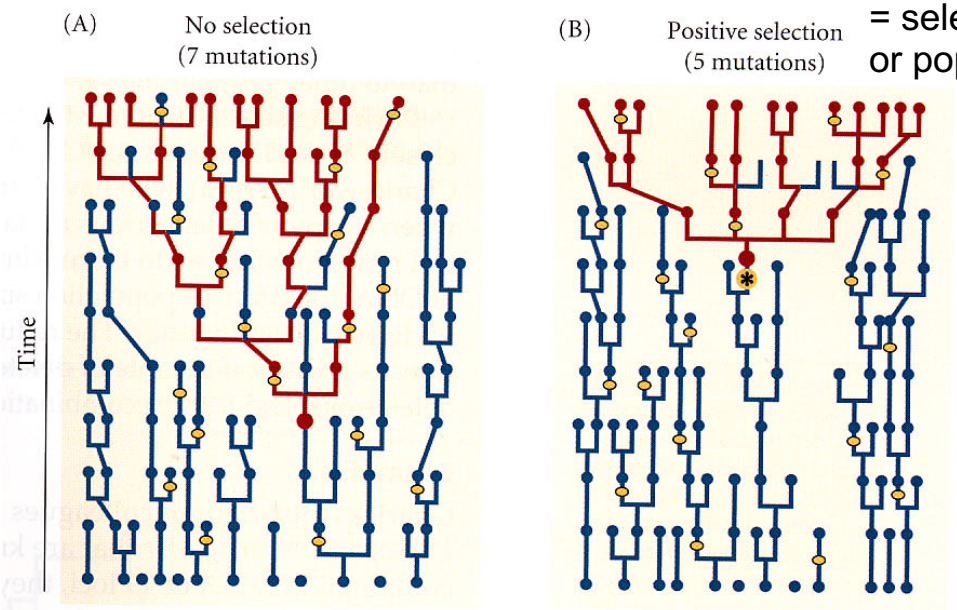
cf. purifying selection (negative selection):
elimination of deleterious alleles from a population

Futuyma, Evolution, Sinauer, 2005

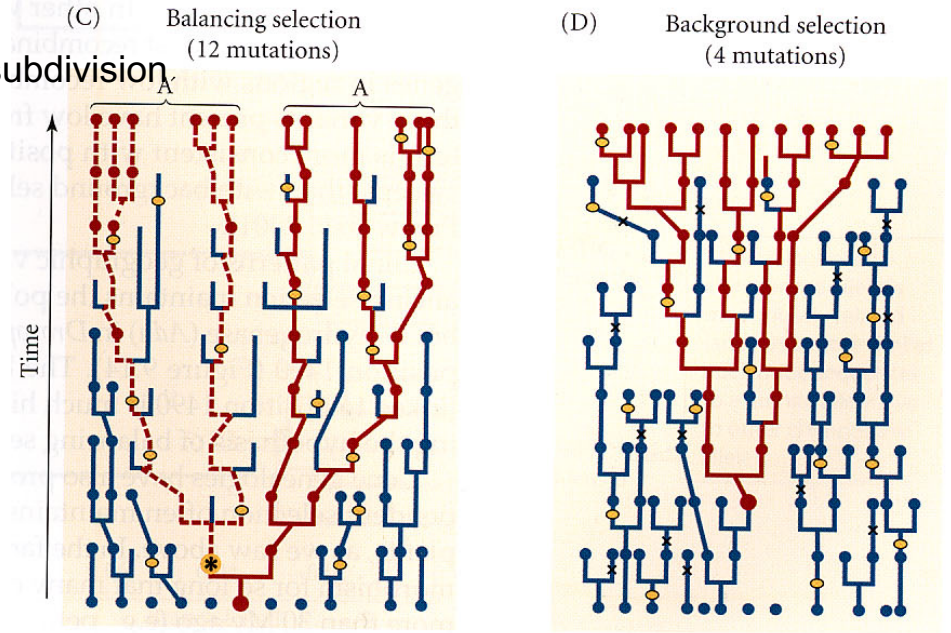
ıew neutral mutations occur among



= selective sweep
or population bottleneck

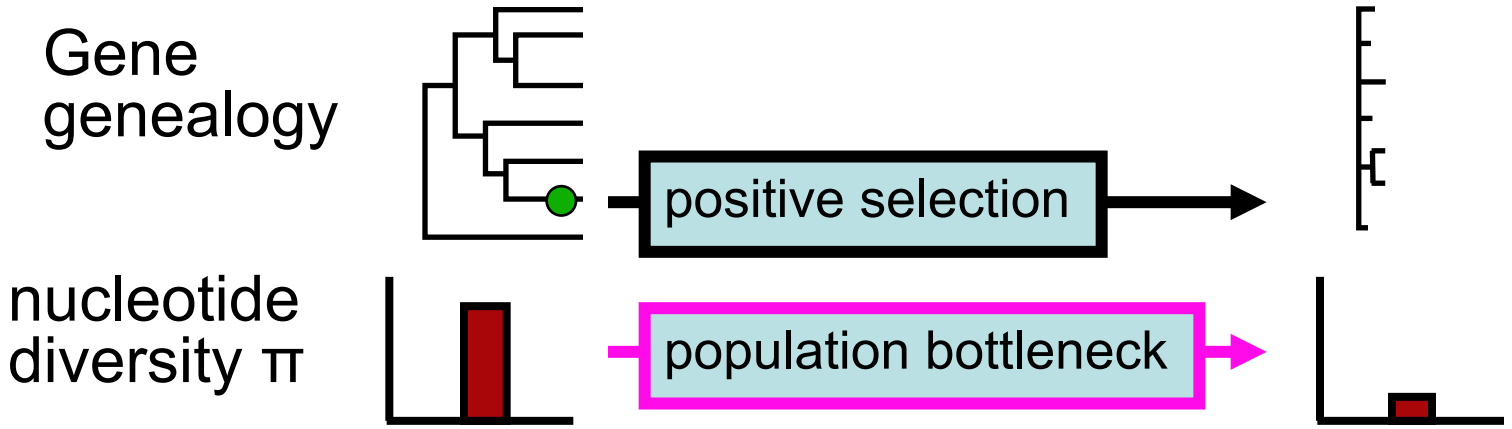


or population subdivision

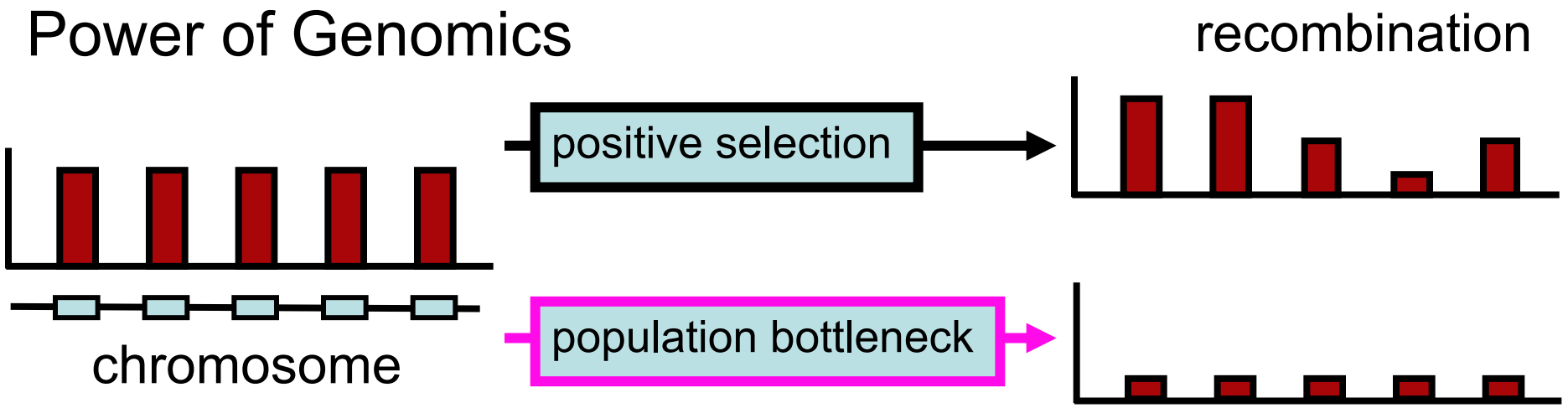


Futuyma, Evolution
Sinauer, 2005, p. 289

Natural selection vs. population processes (or demography)
Natürliche Selektion vs. Demografie

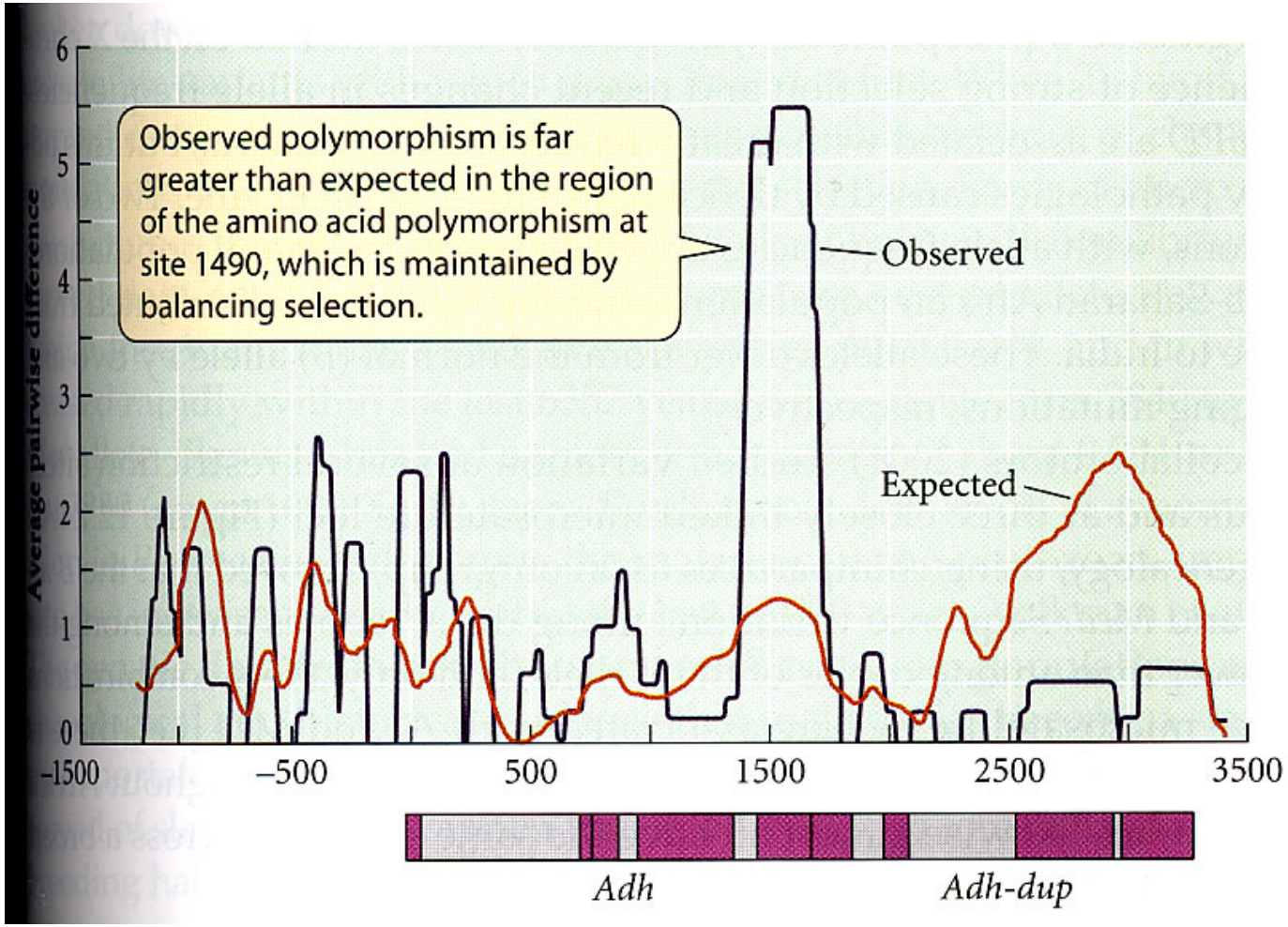


Power of Genomics



π lower than genomic average and neighbors
→ positive selection

balancing selection
("Selektion die Polymorphismen erhält")



Futuyma, Evolution, Sinauer, p. 291

Contents:
molecular population genetics

0. Genome-wide polymorphism and next-generation sequencers

1. nucleotide diversity
selective sweep / hitchhiking, and balancing selection

2. Tajima's *D*

3. positive selection on replacement mutations

θ, another index
(θ, ein weiterer Index)

(nucleotide polymorphism based on polymorphic site / Nukleotid Polymorphismus basierend auf der Anzahl voneinander verschiedener Nukleotide an einer bestimmten Stelle in einem Sequenzvergleich)

```
                        .*....*.*.
(1) AGGCTGCATC          AGGCTGCATC
(2) .A.........    =    AAGCTGCATC
(3) ......T.C.          AGGCTGTACC
```

	(1)	(2)	(3)
(1)	---		
(2)	---	---	
(3)	---	---	---

$$\pi = \sum_{i<j} \pi_{ij} / n_c = ?$$

$$\theta = s / \sum_{k=1}^{n-1} \frac{1}{k} = ?$$

s: the proportion of polymorphic sites (or nucleotide polymorphism) observed in the sample (Das Verhältnis von “voneinander verschieden Stellen” bzw. Nukleotid Polymorphismen, die in einem Sequenzvergleich beobachtet werden können.)

Test of neutrality: Tajima's *D*
(Testen auf Neutralität: Tajima's *D*)

example: balancing selection

Tajima Genetics 123, p. 229, 1989

(1) AGGCTGCATC
(2) ....
(3) .A......C.
(4) .A......C.

$$\pi = \sum_{i<j} \pi_{ij} / n_c = ?$$

$$\theta = s / \sum_{k=1}^{n-1} \frac{1}{k} = ?$$

$$D = \frac{\pi - \theta}{standard\_deviation\_of\_(\pi - \theta)}$$

positive or negative?

Test of neutrality: Tajima's *D*

example: positive selection

(1) AGGCTGCATC

(2) .....

(3) .....

(4) .A.....C.

$$\pi = \sum_{i<j} \pi_{ij} / n_c = ?$$

$$\theta = s / \sum_{k=1}^{n-1} \frac{1}{k} = ?$$

$$D = \frac{\pi - \theta}{standard\_deviation\_of\_(\pi - \theta)}$$

positive or negative?

θ is less affected by frequency

example: balancing selection

(1) AGGCTGCATC
(2) .....
(3) .A.....C.
(4) .A.....C.

example: positive selection

(1) AGGCTGCATC
(2) .....
(3) .....
(4) .A.....C.

$$\theta = s / \sum_{k=1}^{n-1} \frac{1}{k} =$$

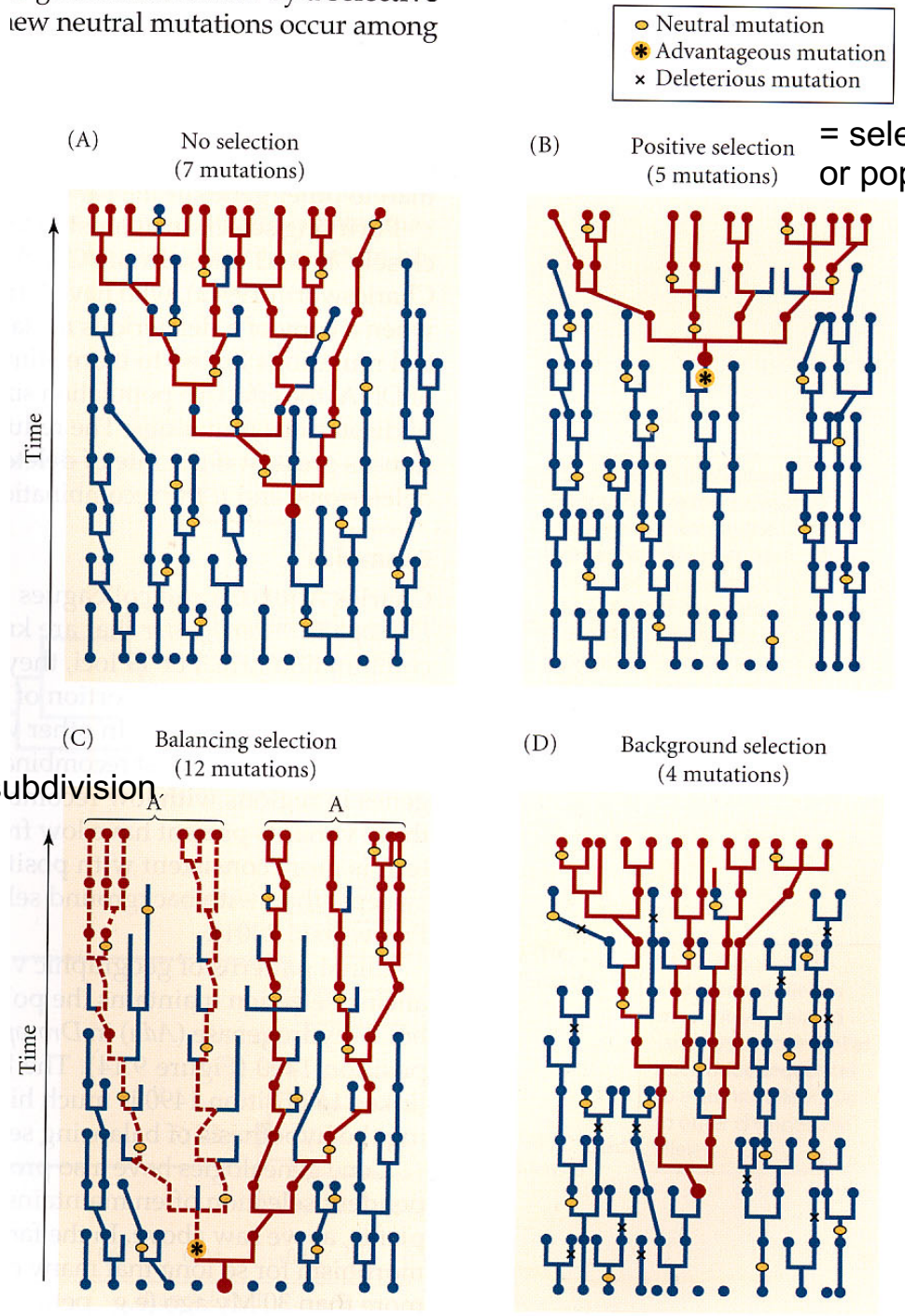
$$D = \frac{\pi - \theta}{standard\_deviation\_of\_(\pi - \theta)}$$

Less than zero when π is low due to positive selection

Intuitively: singleton vs. shared mutation

DnaSP and other free softwares

ıew neutral mutations occur among



= selective sweep
or population expansion

D < 0
more
singletons

or population subdivision

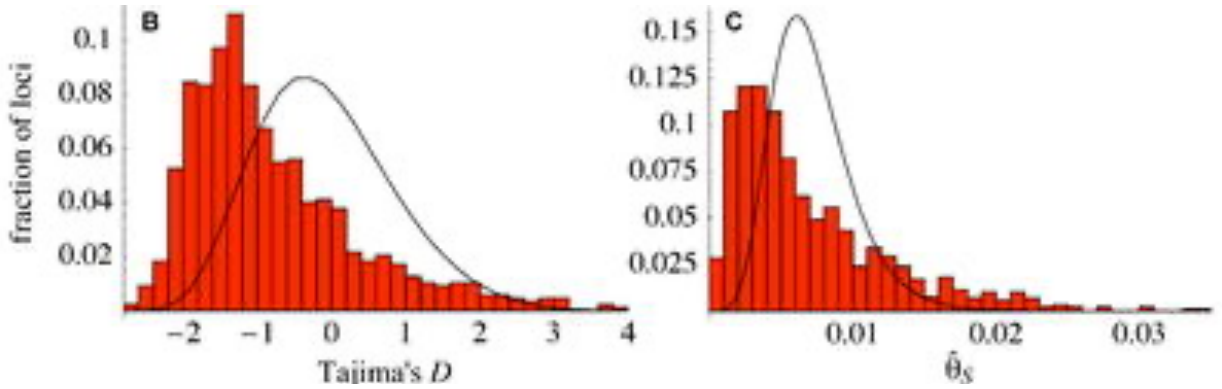
D > 0
fewer
singletons

Futuyma, Evolution,
Sinauer, p. 289

empirical distribution: average may not be zero

876 short fragments in a sample of 96 individuals of *Arabidopsis thaliana*

Nordborg et al. The pattern of polymorphism in *Arabidopsis thaliana*. PLoS Biol. 2005 Jul;3(7):e196



Contents:
molecular population genetics

0. Genome-wide polymorphism and next-generation sequencers

1. nucleotide diversity
selective sweep / hitchhiking, and balancing selection

2. Tajima's *D*

3. positive selection on replacement mutations

Synonymous (silent-site) and nonsynonymous (replacement) mutations

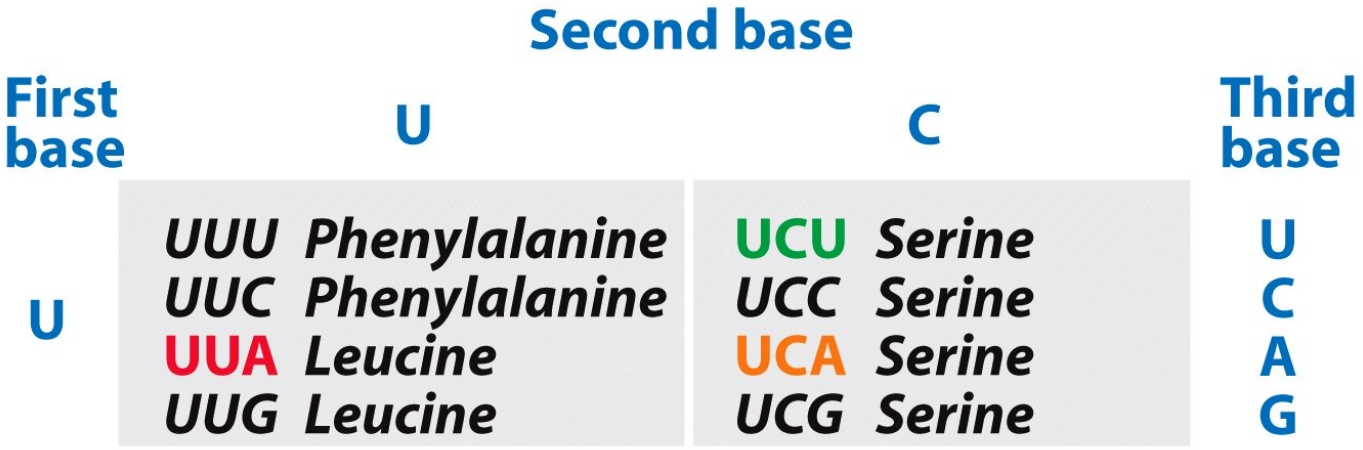


Figure 7-21a Evolutionary Analysis, 4/e
© 2007 Pearson Prentice Hall, Inc.

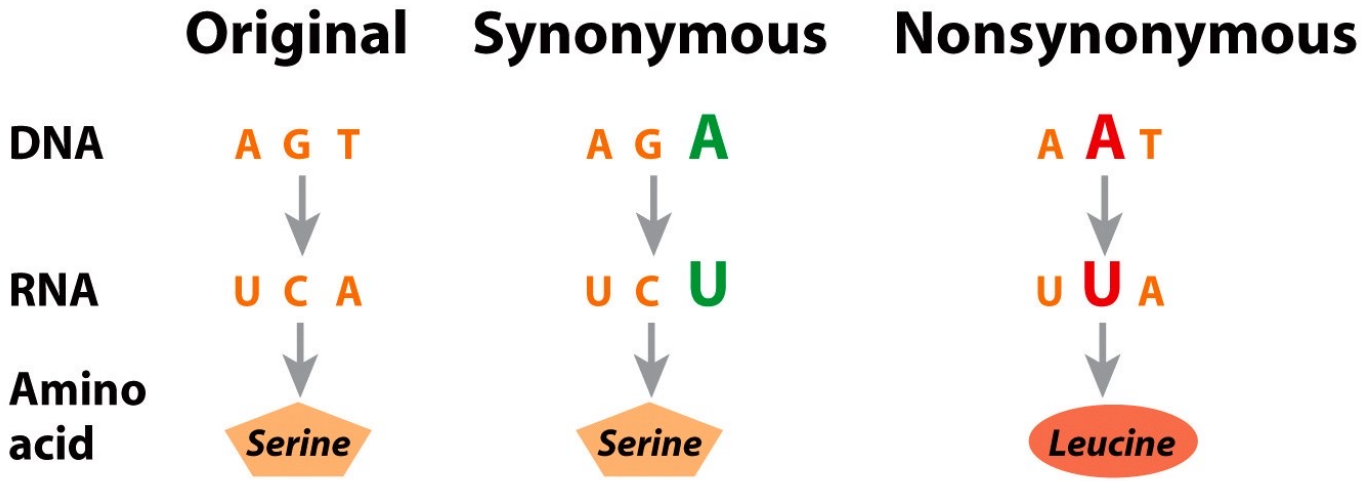


Figure 7-21b Evolutionary Analysis, 4/e
© 2007 Pearson Prentice Hall, Inc.

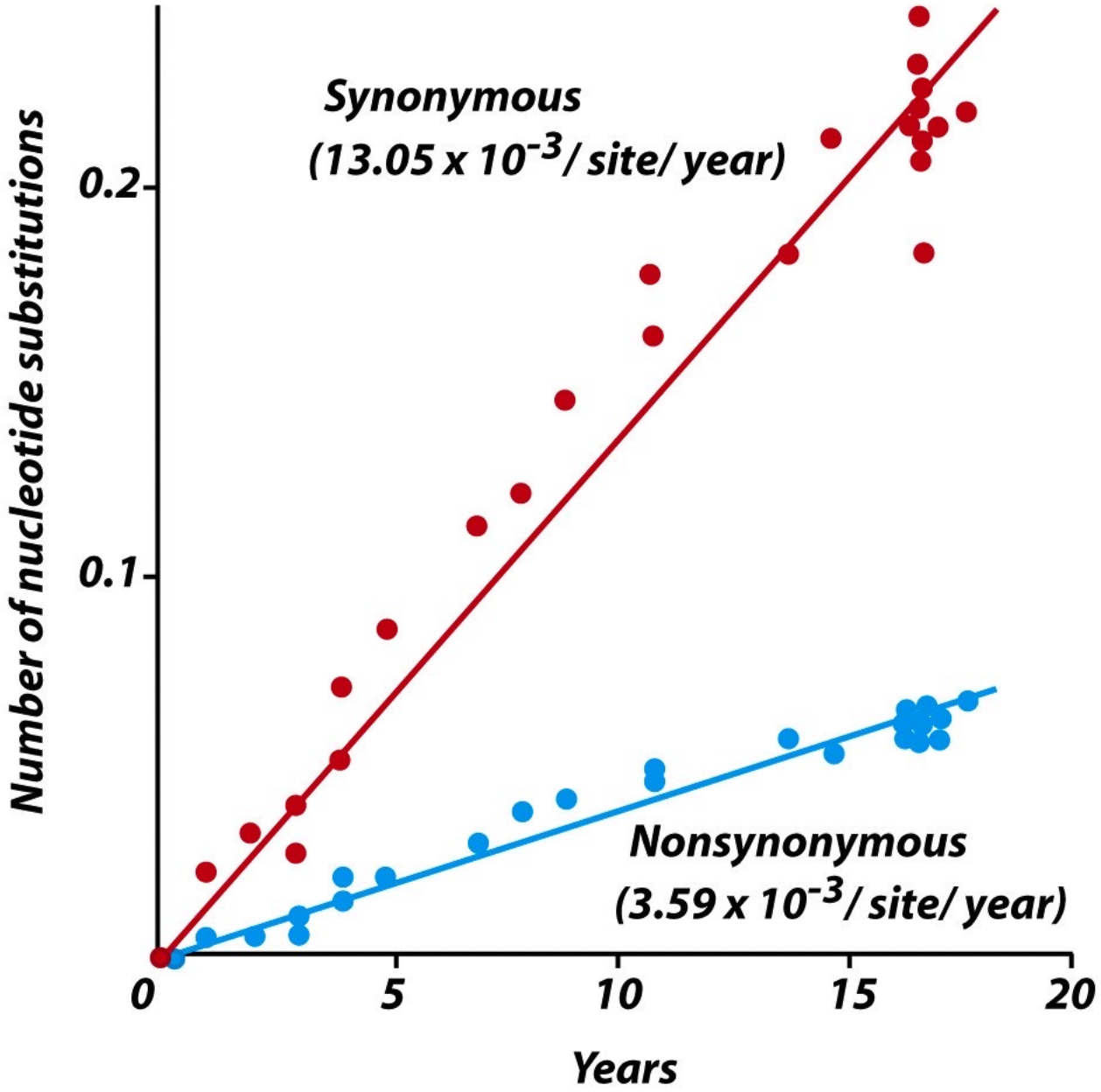


Figure 7-21c Evolutionary Analysis, 4/e

$\omega = K_a/K_s = d_N / d_S$

$d_N / d_S < 1$ when replacements are deleterious

$d_N / d_S = 1$ when replacements are neutral

$d_N / d_S > 1$ when replacements are advangageous (positive selection)

NOTE: $d_N / d_S > 1$ indicates a strong evidence for natural selection, but its sensitivity is low
Sharp (1997) points out that “it is extremely conservative.”

Only when many amino acids changed, it can be more than 1.

d_N / d_S is often used
if only one sequence from each species are available

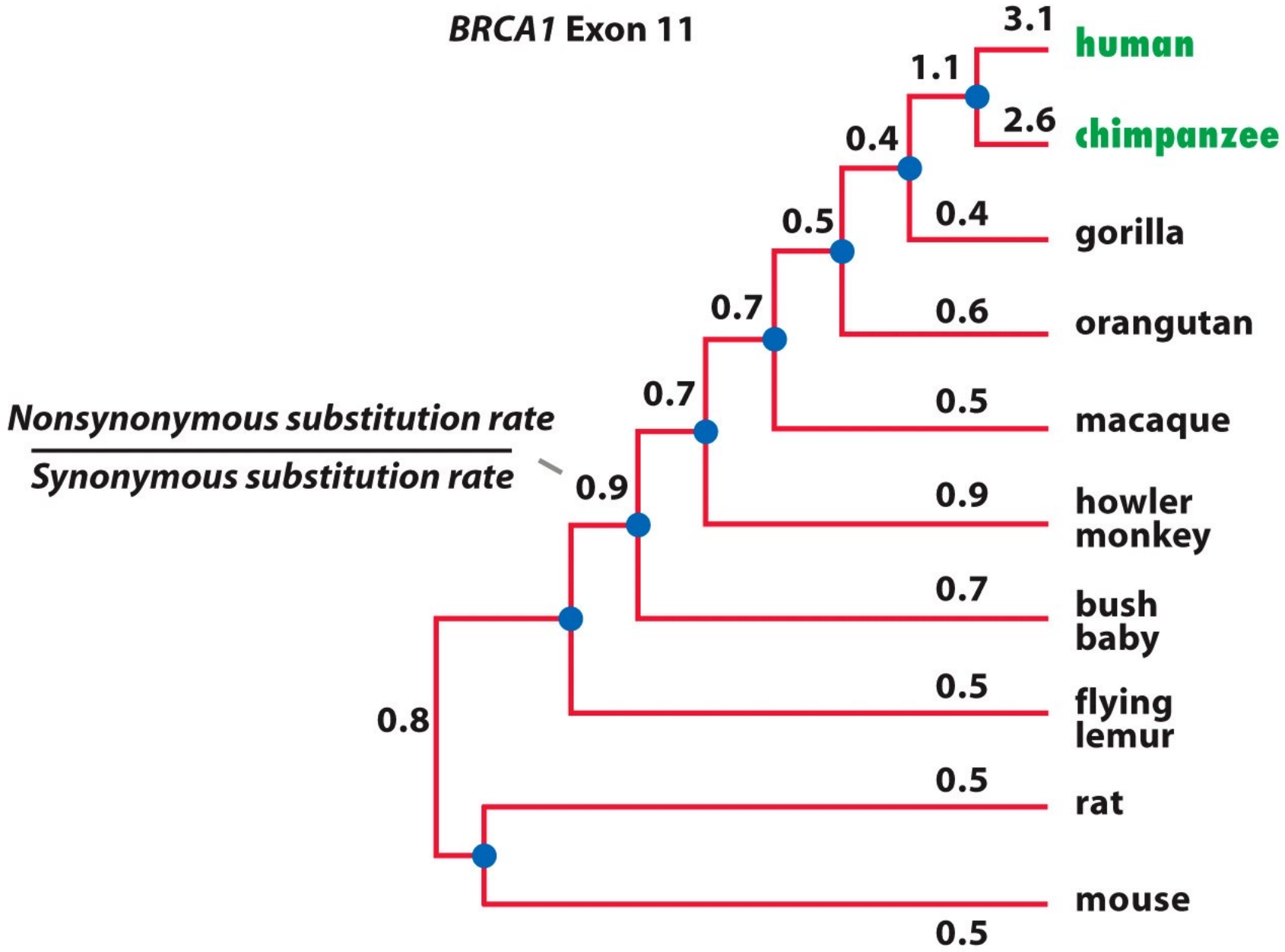


Figure 7-23 Evolutionary Analysis, 4/e

Table 7.2 Studies that confirm positive selection on replacement mutations

Although this table lists just a few examples, it underscores a general point: Evidence for positive selection is particularly strong in genes that code for proteins involved in disease resistance, reproductive conflict, interactions between symbionts, and the development of novel traits.

Gene	Species	Rationale	Reference
MHC Class II	Humans	Strong selection for divergence among antigen-recognition proteins	Hughes, A.L., and M. Nei. 1989. *Proceedings of the National Academy of Sciences USA* 86: 958–962.
Semenogelin II gene	Primates	Strong selection for semen coagulation in species with intense sperm competition	Dorus, S. D., et al. 2004. *Nature Genetics* 12: 1326-1329.
Lysin protein and receptor	Abalone	Strong selection on species-specific egg-recognition proteins on sperm	Swanson, W.J, and V.D. Vacquier. 1998. *Science* 281: 710–712.
Eosinophil cationic protein	Primates	Strong selection on a recently duplicated gene involved in disease resistance	Zhang, J., H.F. Rosenberg, and M. Nei, 1998. *Proceedings of the National Academy of Sciences USA* 95: 3708–3713.
Self-incompatibility loci	Tomato family plants	Strong selection for divergence among proteins involved in self-fertilization	Clark, A.G., and T.-H. Kao. 1991. *Proceedings of the National Academy of Sciences USA* 88: 9823–9827.
Multicolored fluorescent proteins	Reef-building corals	Evolving algal sybionts impose selection on ability of host corals to regulate them.	Field, S. F., et al. 2006. *Journal of Molecular Evolution* 62: 332-339.
ASPM	Humans and great apes	Selection for increased brain size.	Evans, P. D., et al. 2004. *Human Molecular Genetics* 13: 489-494.

Table 7-2 Evolutionary Analysis, 4/e

Case study: Evolution of language

Was there positive selection to have language?
How many genes were necessary to obtain language?
How and when did the language evolve?

Did Neanderthal have language,
which diverged 3-400,000 years ago?

Limited evidence: archaeology, fossil of skeletons

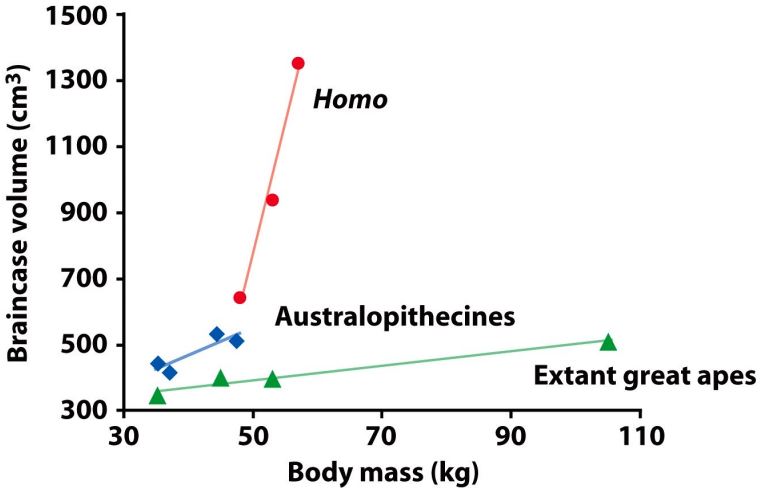


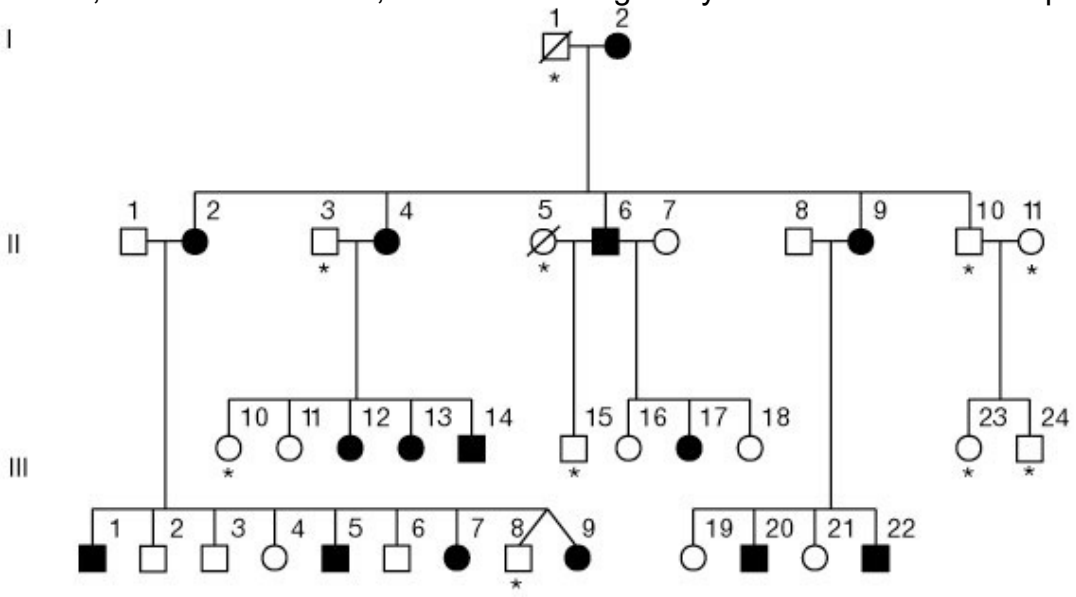
Figure 20-32 Evolutionary Analysis, 4/e
© 2007 Pearson Prentice Hall, Inc.

Figure 20-31 Evolutionary Analysis, 4/e
© 2007 Pearson Prentice Hall, Inc.

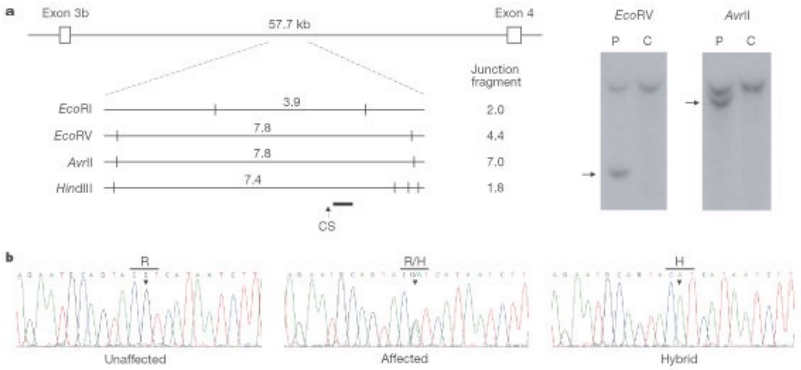
From Freeman and Herron (2007) "*Evolutionary Analysis (4th edn)*"

Candidate gene from human disease genes: FoxP2, necessary for the development of Broca's area and language

FIGURE 1. Pedigree of the KE family.
Affected individuals are indicated by filled symbols. Asterisks indicate those individuals who were unavailable for genetic analyses. Squares are males, circles are females, and a line through a symbol indicates that the person is deceased.



A forkhead-domain gene is mutated in a severe speech and language disorder
Lai et al. *Nature* **413**, 519-523, 2001



The signatures of recent positive selection

(Anzeichen von rezenter positiver Selektion)

Table 1 **Variation at the *FOXP2* locus in humans**

No. of chromosomes sequenced	40
Length covered (double stranded, all individuals)	14,063 bp
Divergence from the chimp sequence*	0.87%
No. of variable positions	47
Singletons (no. of variable sites occurring at frequency 1 and 39)	31
θ_W (nucleotide diversity based on the no. of polymorphic sites)	0.079%
θ_π (mean nucleotide diversity)	0.03%
θ_H (nucleotide diversity with more weight given to alleles at high frequency[17])	0.117%
D ($P < 0.01$)†	−2.20
H ($P < 0.05$)‡	−12.24

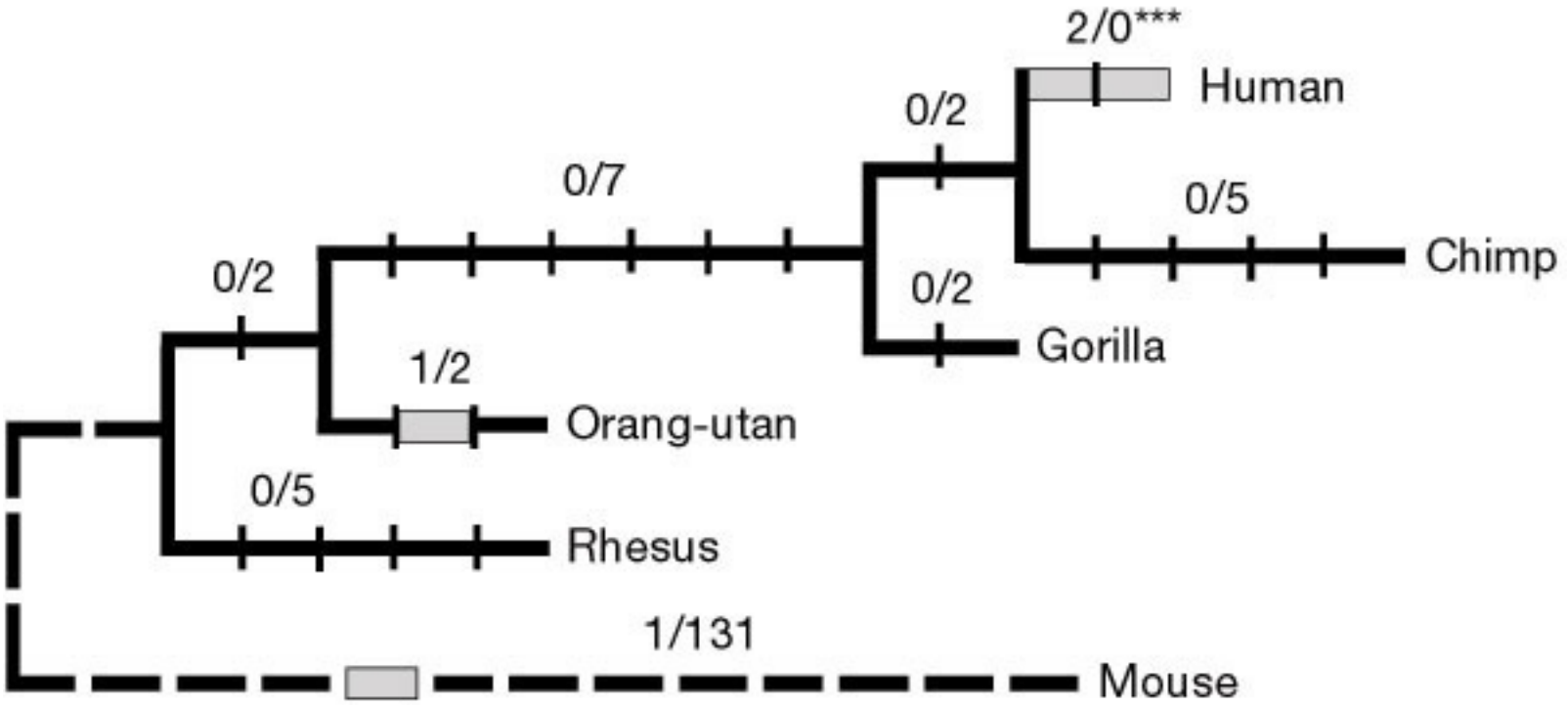
π →

*The corresponding value for the orang-utan is 2.5.
†A negative D value indicates a relative excess of low-frequency alleles[15].
‡A negative H value indicates a relative excess of high-frequency derived alleles[17].

Molecular evolution of *FOXP2*, a gene involved in speech and language
Wolfgang Enard, Molly Przeworski, Simon E. Fisher, Cecilia S. L. Lai, Victor Wiebe, Takashi Kitano, Anthony P. Monaco and Svante Pääbo
Nature **418**, 869-872, 2002

Zheng et al., Genetics 162, p. 1825 (2002)

high Ka/Ks in the human lineage
(hoher Ka/Ks in der menschlichen Abstammung)



When did the language appear?

Krause et al. (2007) Curr. Biol. 17: 1908

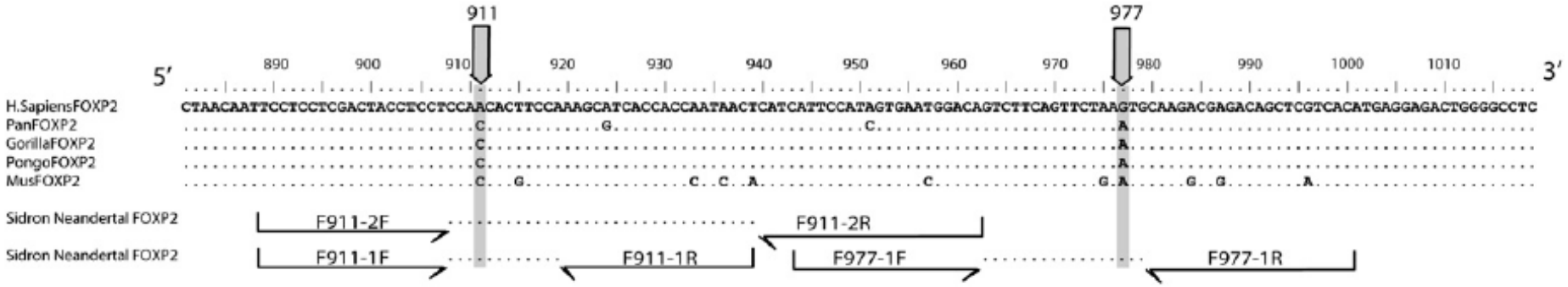


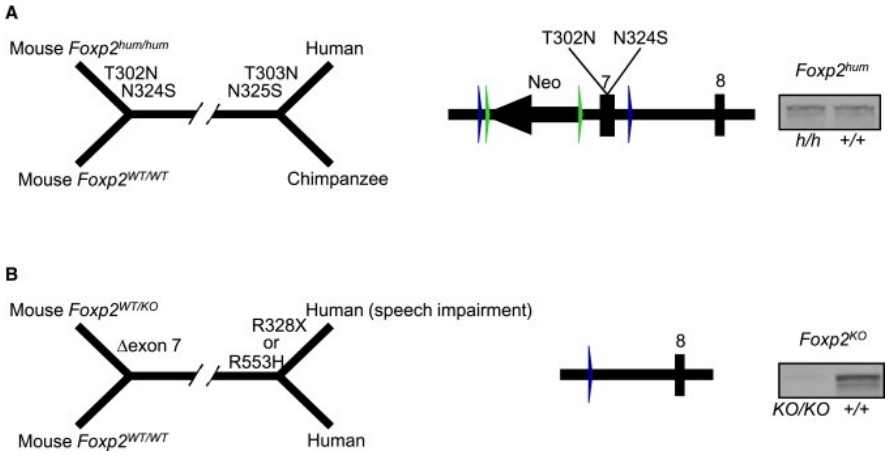
Figure 1. Sequence Alignment of Nucleotide Positions 880–1020 from the *FOXP2* Gene

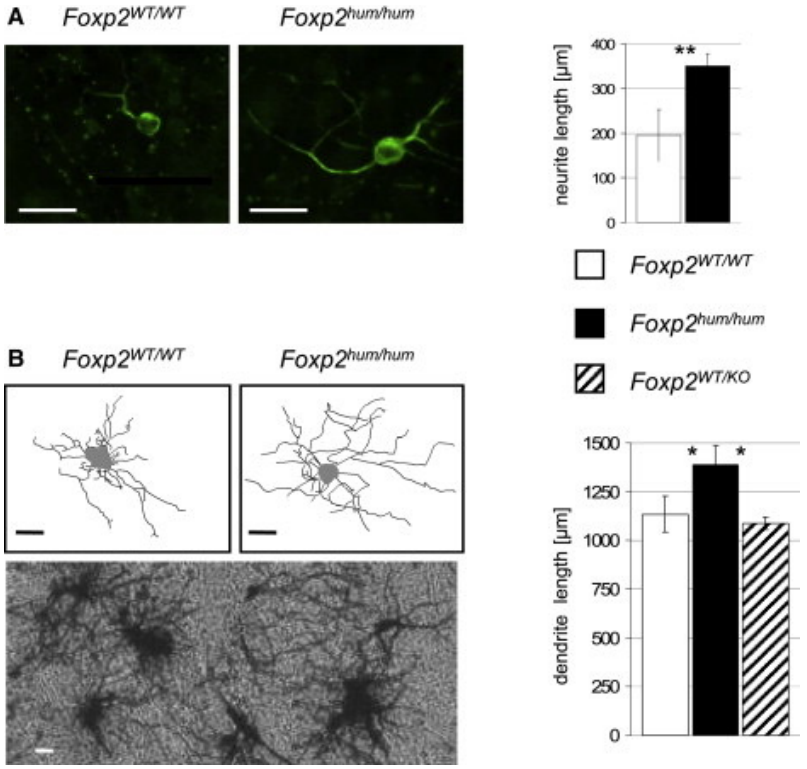
The two nonsynonymous nucleotide substitutions on the human lineage are indicated by arrows. Identical positions in the alignment are given as dots. The three primer pairs used to retrieve the two substitutions from the El Sidrón Neandertals are indicated by arrows.

Time estimate based on coalescent simulation:
with approximate 95% confidence intervals of 0 and 120,000 years ago
--> Time estmiate has a number of assumptions and may not be accurate

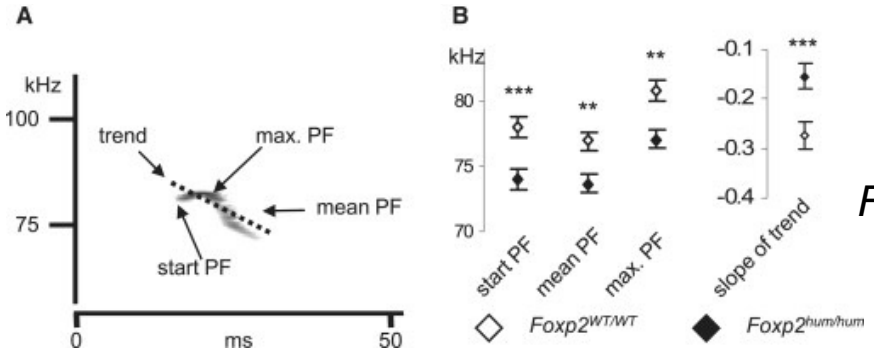
Note: There is little evidence that this gene really contributed to the evolution of language. Even if so, multiple genes must have been necessary.

Transgenic mouse with human FoxP2





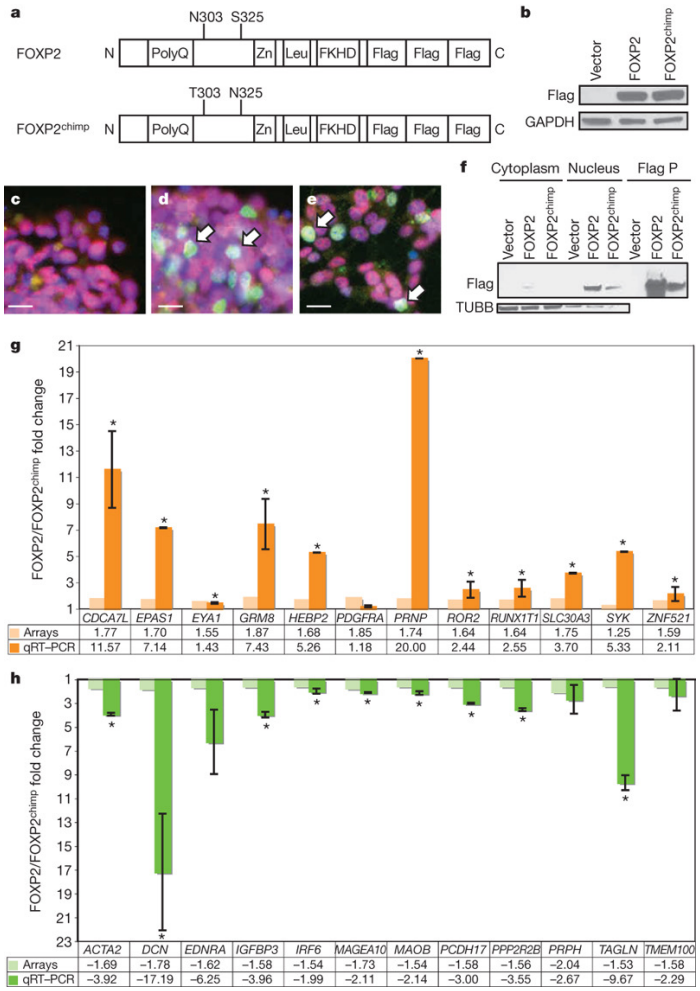
Foxp2^{hum} Increases the Length of Dendritic Trees



Foxp2^{hum} Affects the Structure of Pup Isolation Calls

Enard et al., Cell (2009) 137:961

FOXP2 and FOXP2^{chimp} differentially regulate genes in neural cells.



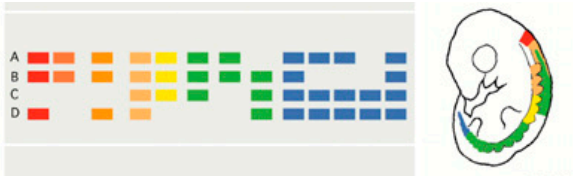
G Konopka *et al. Nature* **462**, 213-217 (2009)
doi:10.1038/nature08549

Summary

- What was revealed about evolution by using molecular data?
- Why are model species studied in addition to *Homo sapiens*?

Exercise example: polyploid species

- Genome duplication (polyploidization) is prevalent in animals, fungi and plants (>35%)



Ohno 1970, Swalla, *Heredity* 2006

Leitch & Leitch *Science* 2008

A new species during the past 150 years in Switzerland

A textbook example in Switzerland:

(picture by National Geographic)

Urnerboden
past 150 years

drier

Cardamine rivularis (2*x*)
C. pratensis (4x)

intermediate, fluctuating

insueta (3*x*)
schulzii (6x)

wet

amara (2*x*)

Lihova, Shimizu & Marhold *MPE* 2006
Mandakova et al., *Plant Cell* 2013
Lihova et al. *Ann Bot,* 2014

A new species during the past 150 years in Switzerland

https://www3.nhk.or.jp/nhkworld/en/tv/scienceview/ (freely available until Aug 2020)

Challenge in the polyploid transcriptome: quantifying ratio of homologous pairs

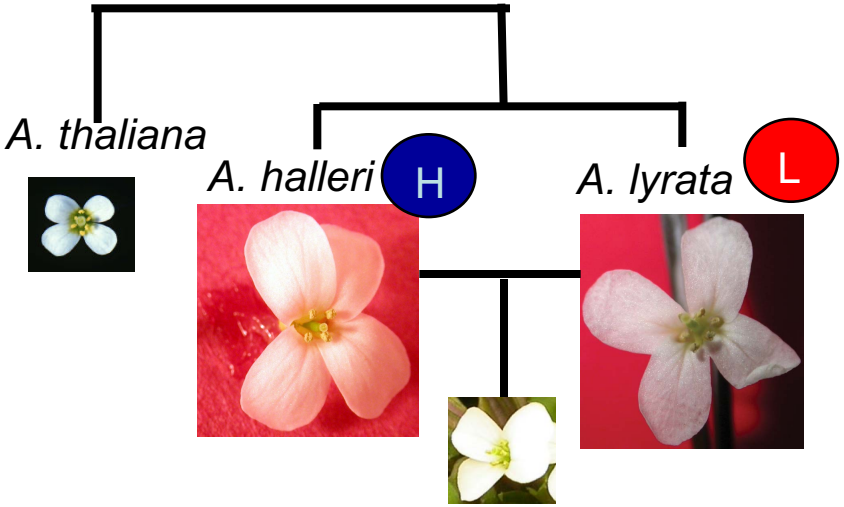
Major obstacle: difficult to separate homeologous pairs due to high homology

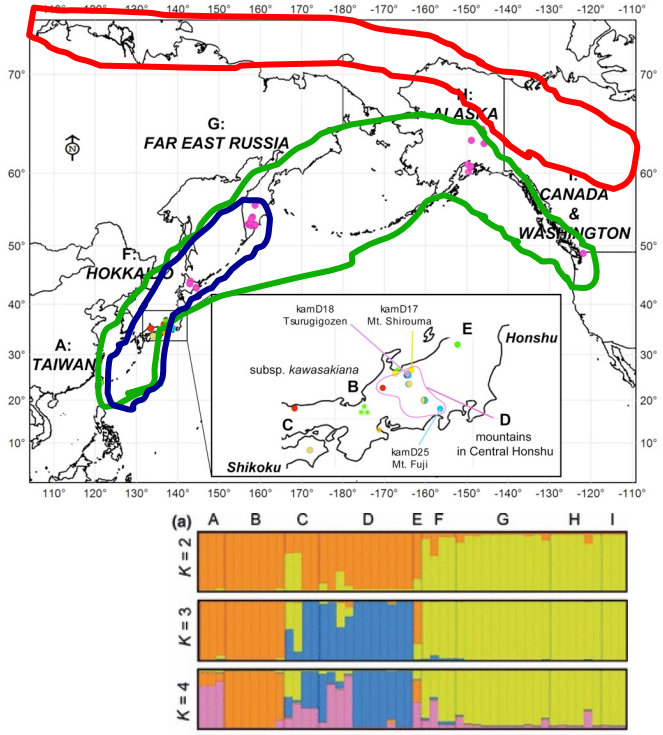
1. Genome assembly of polyploid is difficult
   -> new sequencers are solving it
2. Transcriptome
   - Designing homeolog-specific primer is time-consuming
   - Few tools to deal with genome-wide pattern

(SNP-based method: Shi et al. *Nature Comm* 2012, Page et al. *G3*, 2013)

-> New bioinformatic workflow HomeoRoq

Arabidopsis kamchatica
a broadest ecological niche in the genus *Arabidopsis*



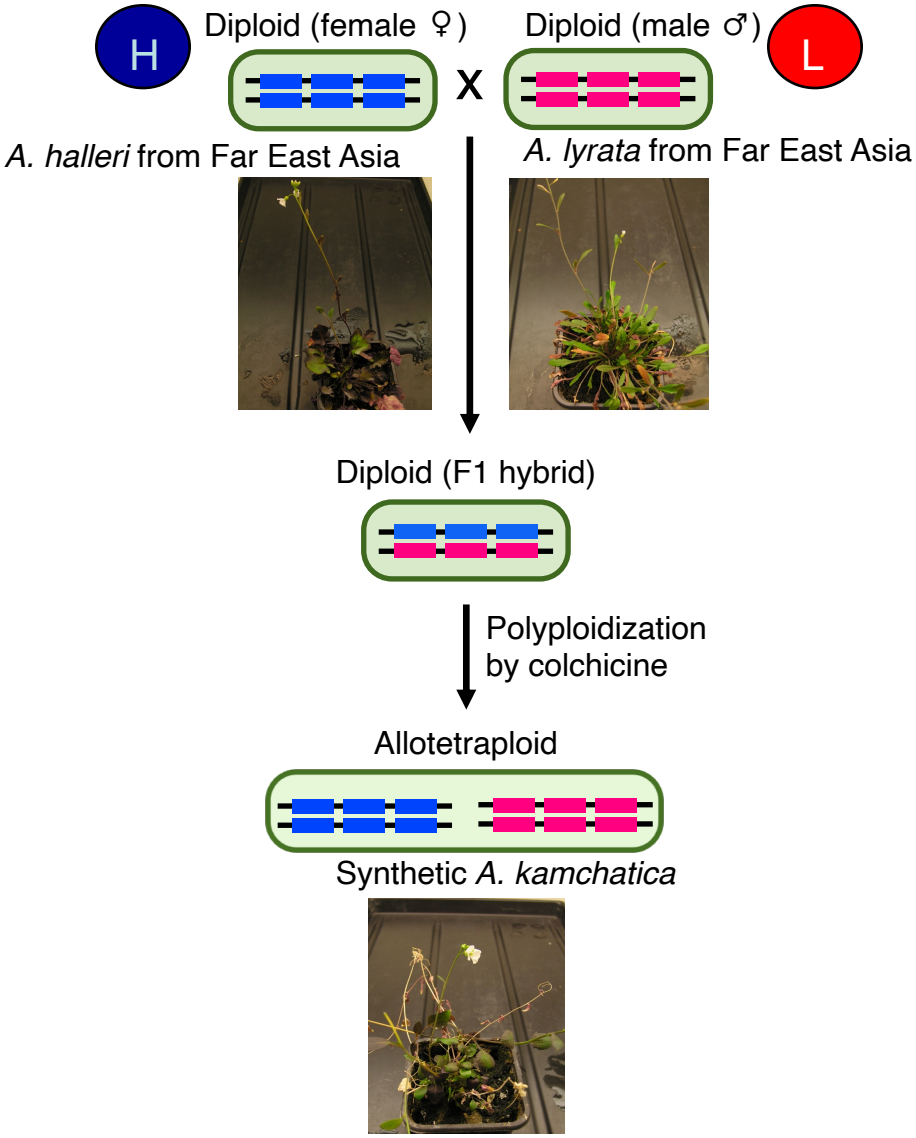


Shimizu-Inatsugi et al. *Mol Ecol* 2009

ssp. *kamchatica,* Perennial
0-3000 m alt

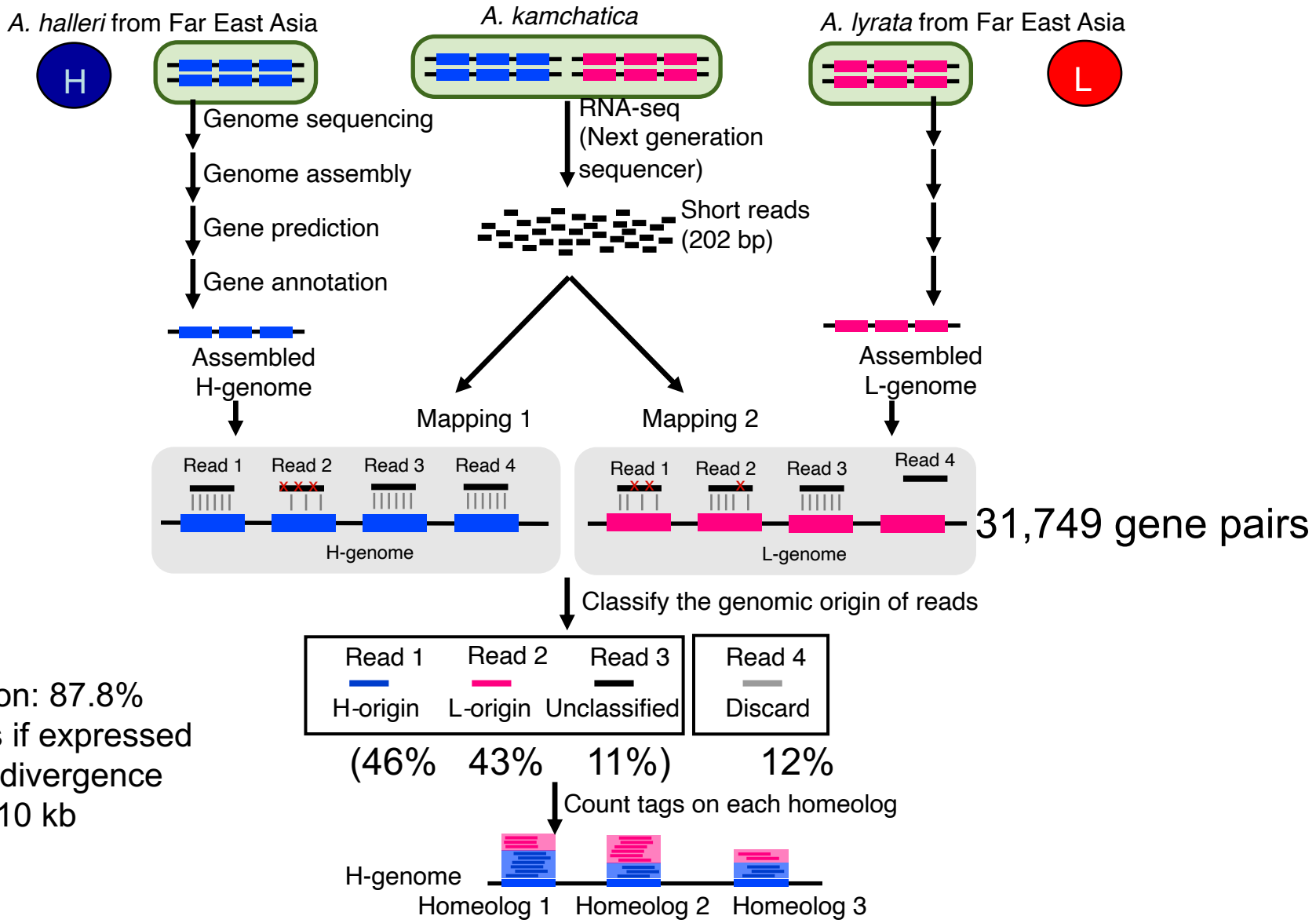
ssp. *kawasakiana*
Annual
hot sea- or lake shore
0-85 m alt in Japan

Simplest polyploid model: Genome assembly of two parental species and synthetic allopolyploid



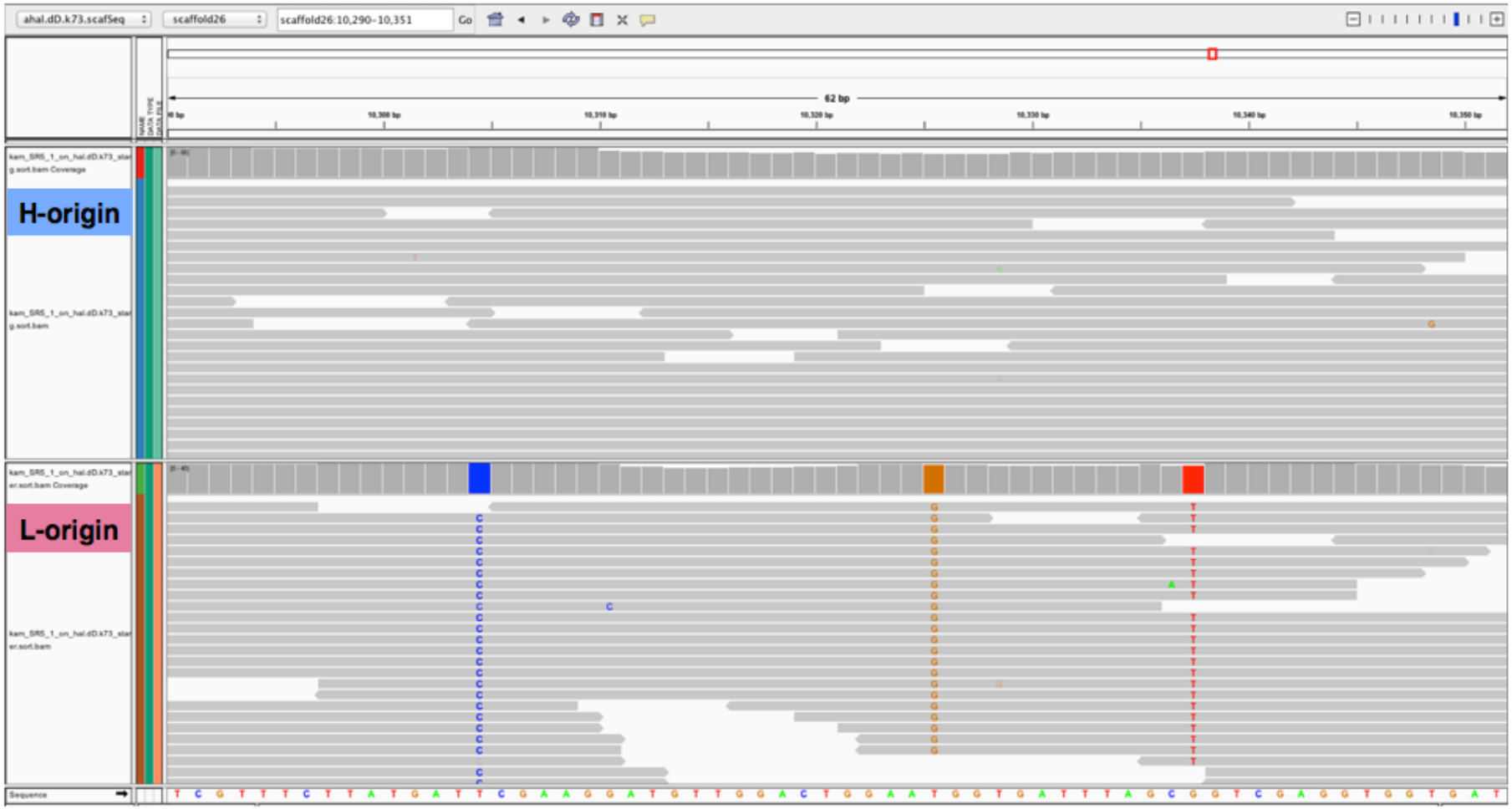
Akama et al. *Nuc Acids Res* 2014

HomeoRoq (Homeolog Ratio and Quantification): bioinformatic workflow using NGS

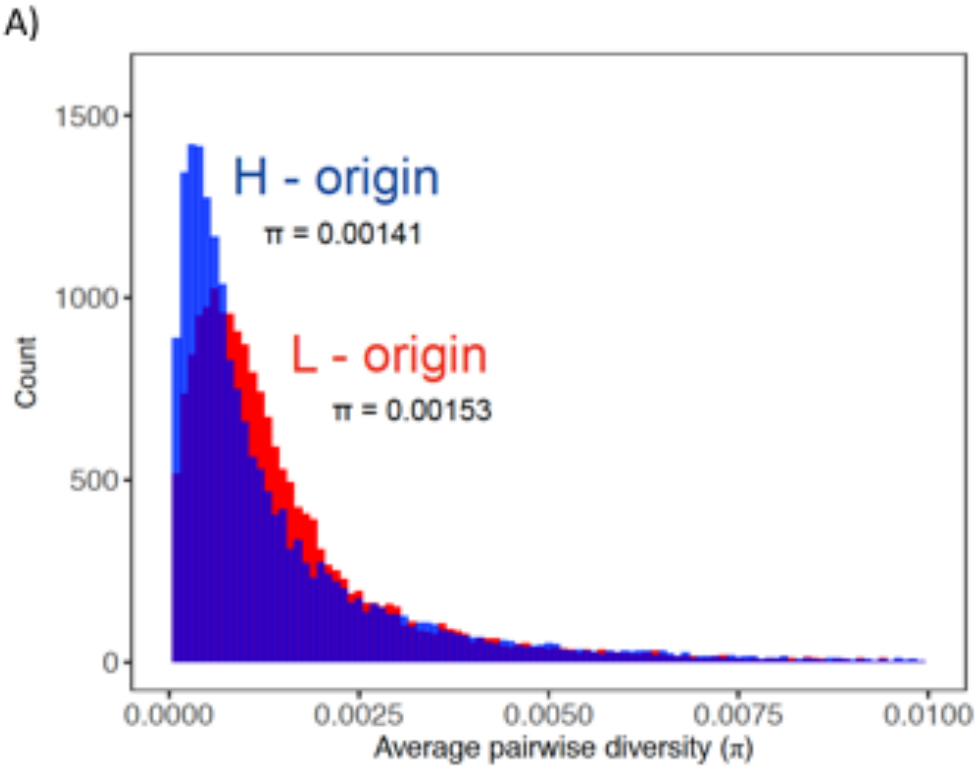


Simulation: 87.8%
of genes if expressed
- 2.44% divergence
- N50 ~ 10 kb

Visualization:
homeologs were well separated



nucleotide diversity of polyploid species



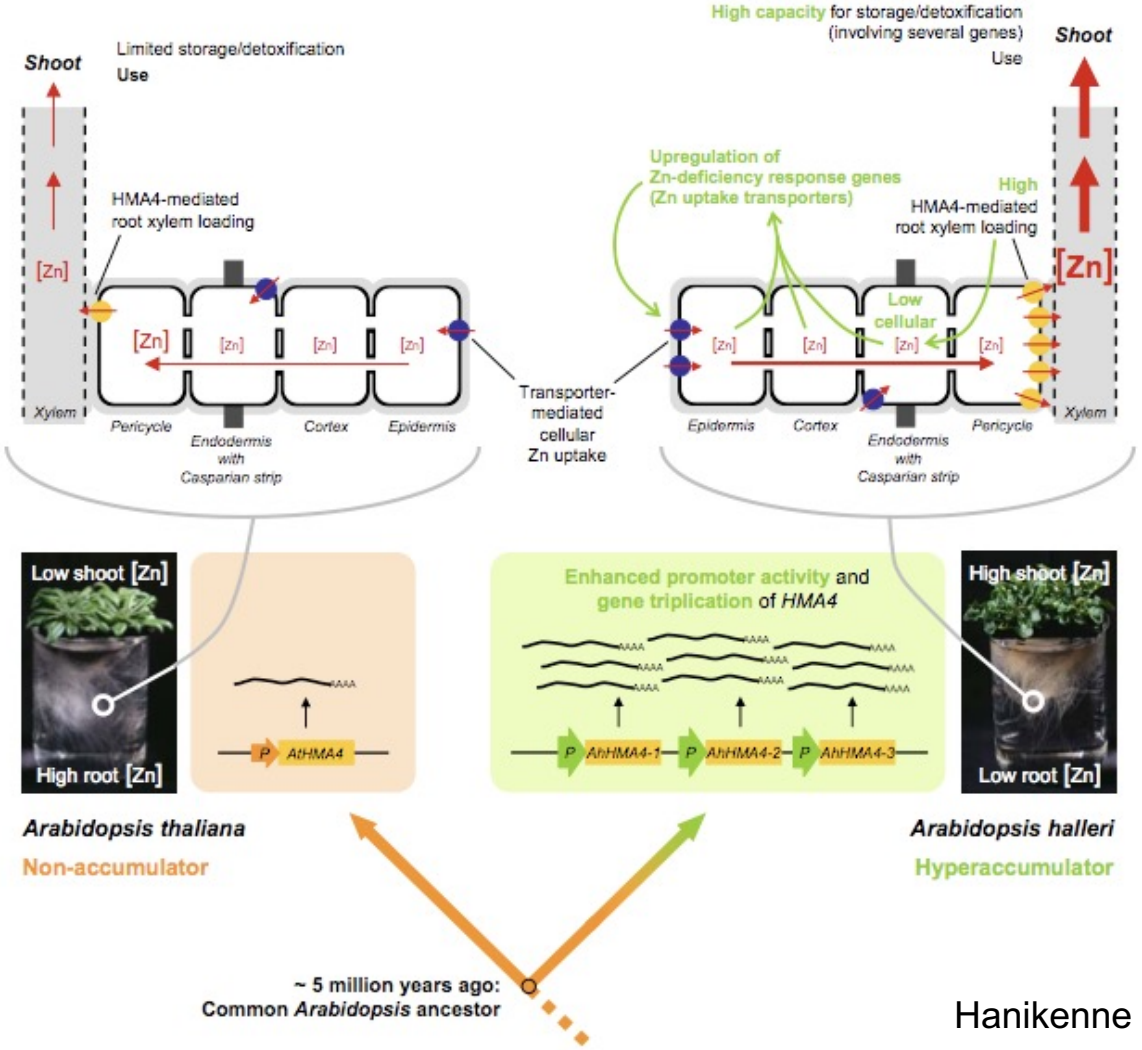
synonymous π: 0.0049 in H-subgenome, 0.0044 in L-subgenome

similar to *A. thaliana* (0.0059 - 0.007)

Several times lower than parental species (0.028 for *A. halleri* and *A. lyrata*) suggesting a bottleneck at the polyploidization events

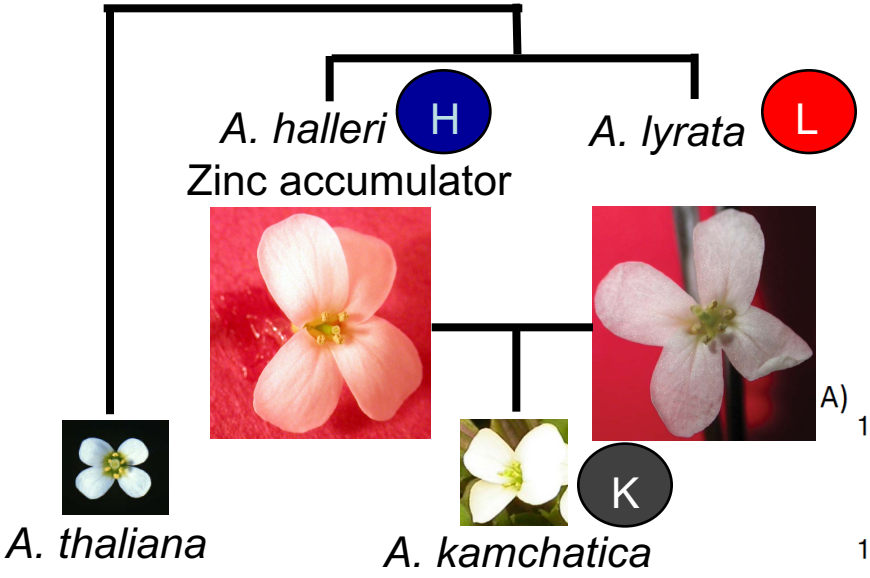
Paape et al. *Nature Comm* 9:3909, 2018

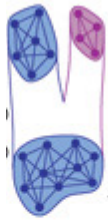
Hyperaccumulation and phytoremediation *HMA4* transporter gene



Hanikenne et al 2008 Nature

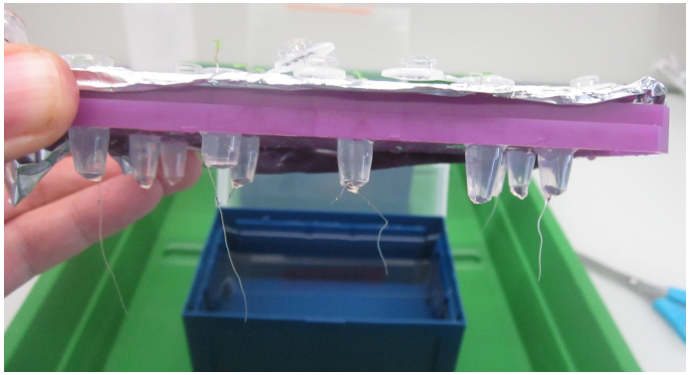
Zinc hyperaccumulation to study both phenotype and gene expresion level quantitatively



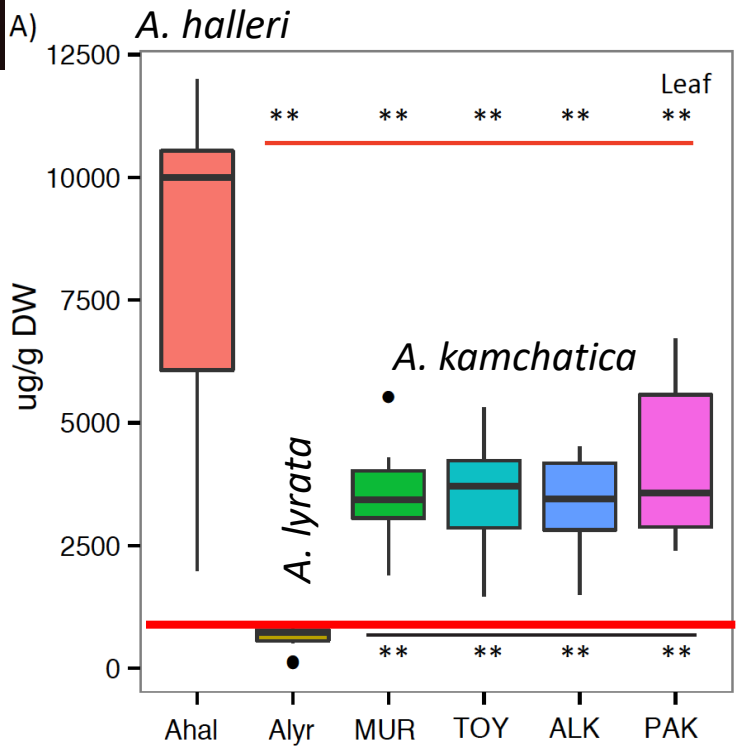


What happens?

Paape et al.,
Mol Biol Evol 2016



500 uM Zinc for 7 days treatment



Insect defense

Soil contamination

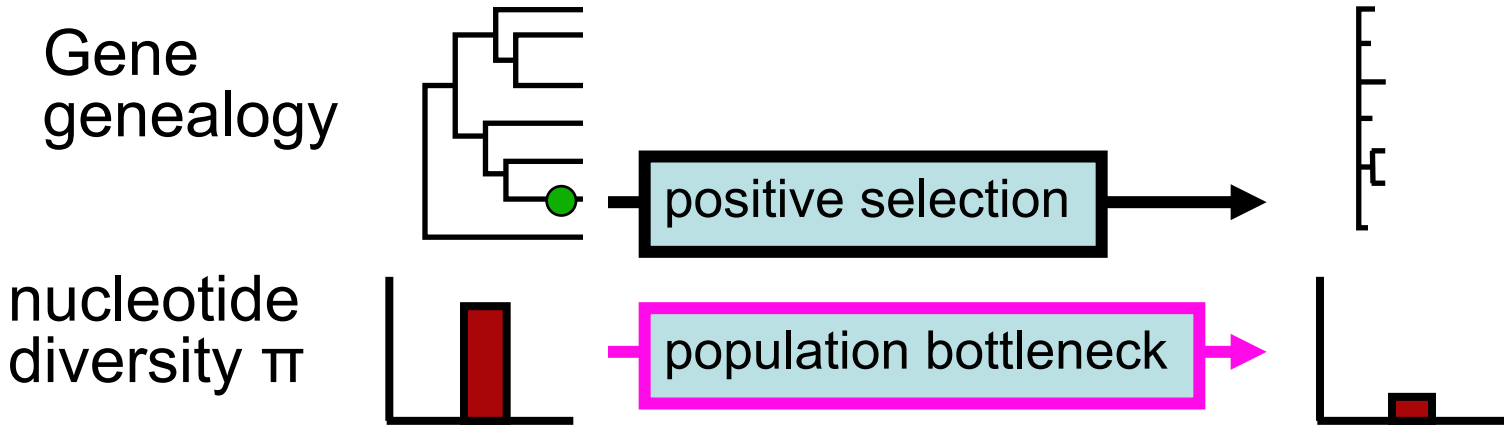
- Silver from Tada mine during the 16th century
- Major source to build the Osaka Castle when Hideyoshi Toyotomi unified Japan
- Closed in 1973

Zinc concentration of habitat soil (mg kg^{-1})

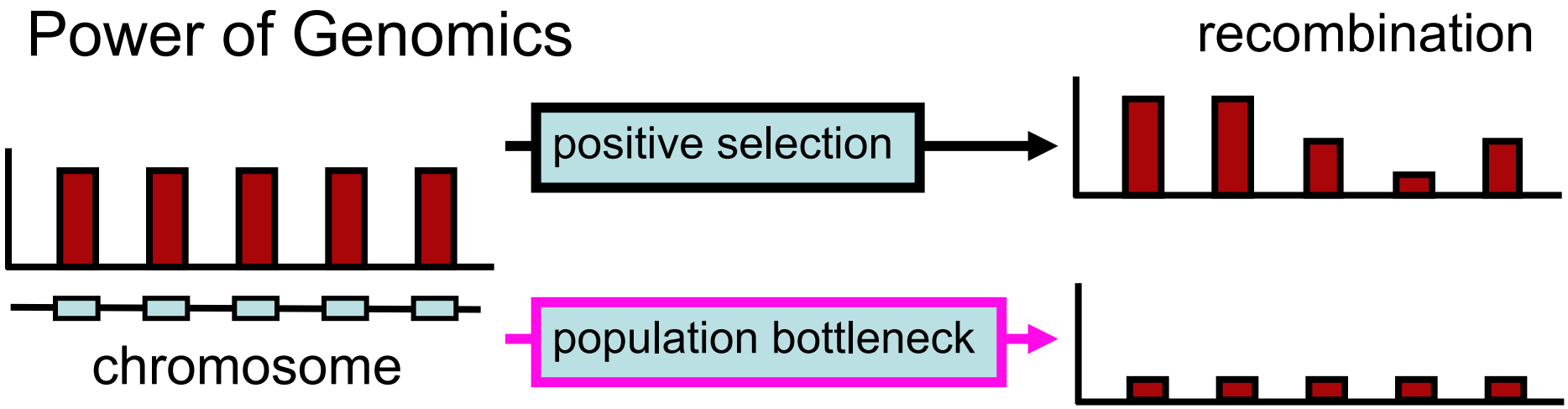
A. halleri:
3.4-20 (nonmetalliferous) to 200-3,200 (metaliferous)
(Hanikenne et al. *PLoS Genet* 2013)

A. kamchatica 30 localities*:* 31-1,100

Natural selection vs. population processes (or demography)
Natürliche Selektion vs. Demografie

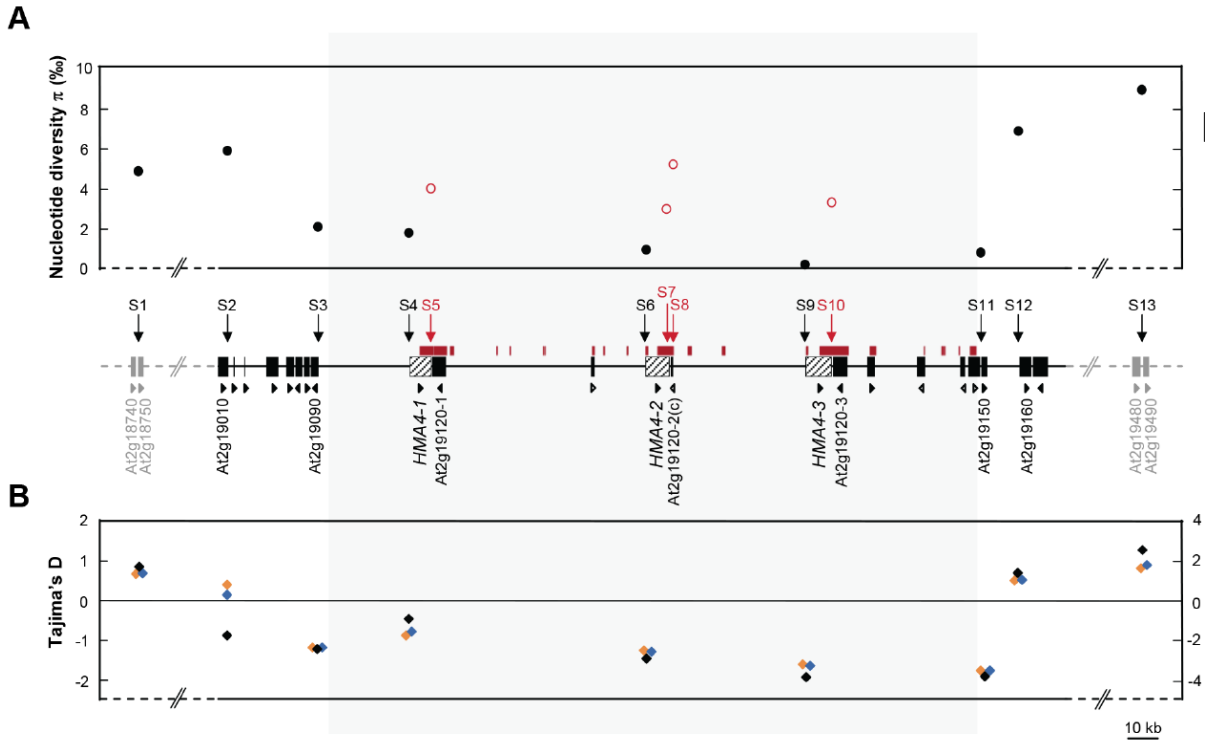


Power of Genomics



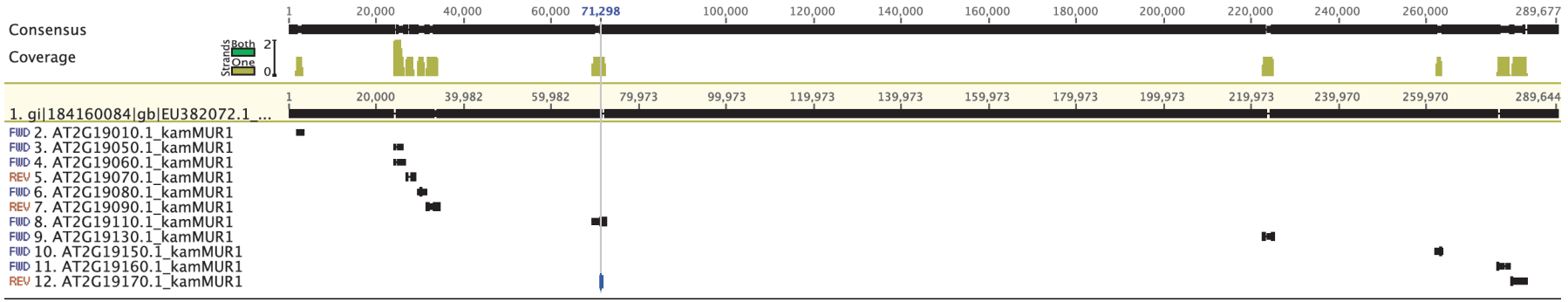
π lower than genomic average and neighbors
→ positive selection

HMA4 diversity statistics in the diploid parent *A. halleri*



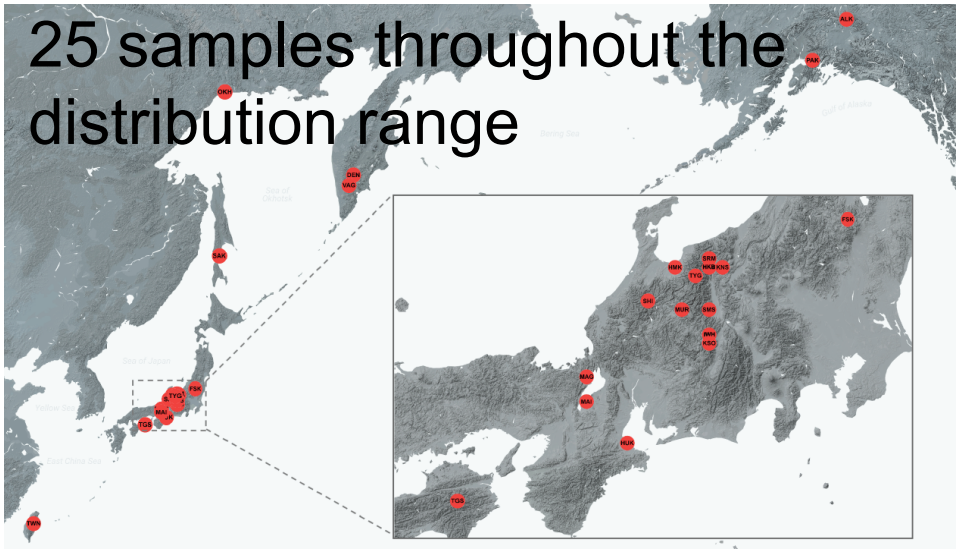
Hanikenne et al. 2013
PLoS Genetics

"Hard" selective sweep

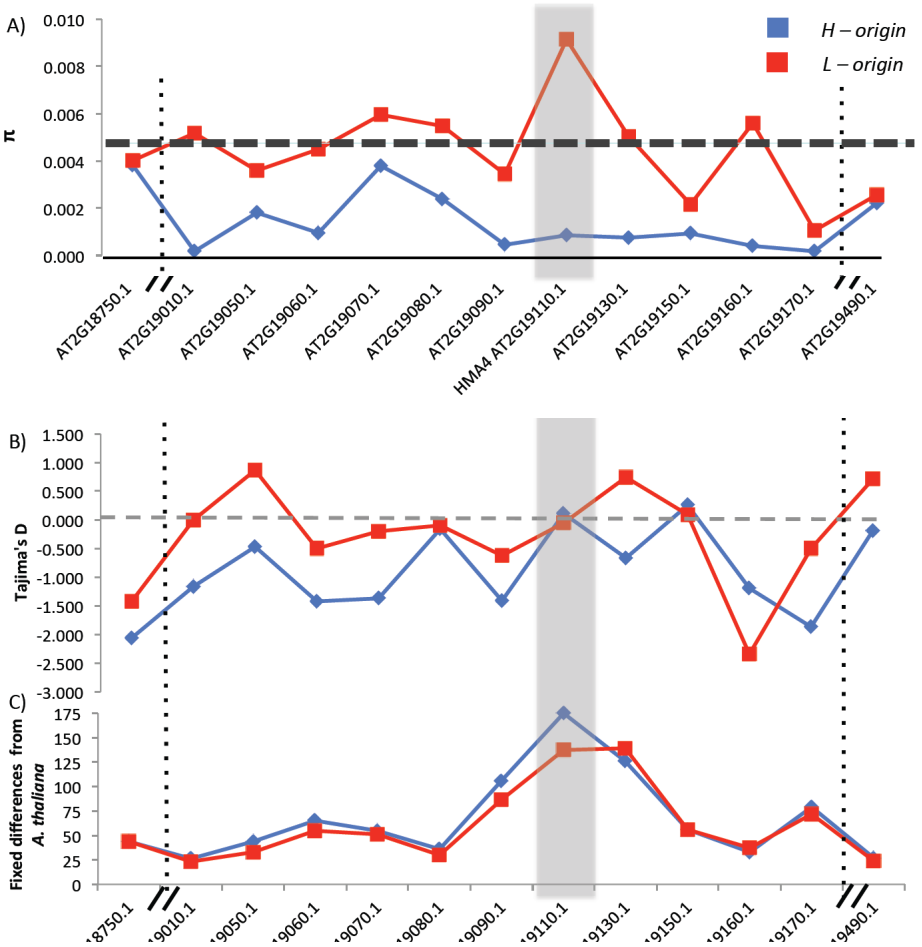


Resequencing of 25 genotypes: Signature of selective sweep at the *HMA4* region

Genome-wide silent nucleotide diversity π = 0.0045 (similar to diploid *A. thaliana*)

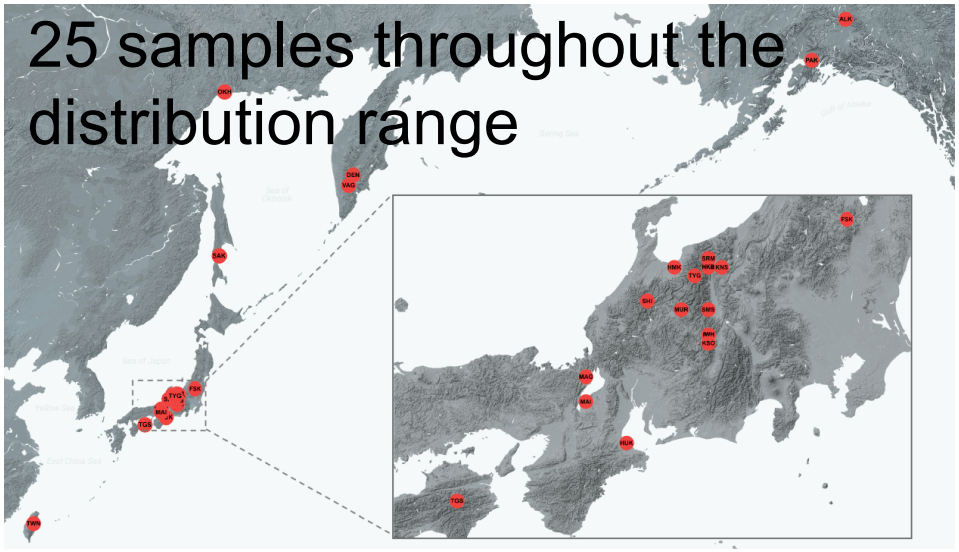


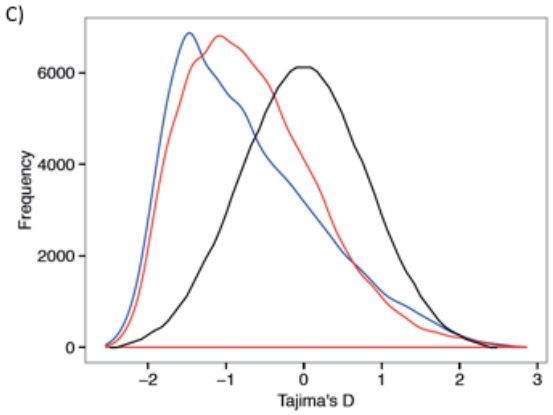
Inherited low π in *halleri*-derived *HMA4* regions



Paape et al., *Mol Biol Evol* 2016

Different signature of selection





Synonymous nucleotide diversity: correlation is low (r = 0.04)

Tajima's *D*: Low correlation (r = 0.05)

Genome-wide pattern: different signatures of selection

Increased number of selection targets

Paape et al. *Nature Comm* 9:3909, 2018

How about other allopolyploid species?
Rapid advance in polyploid genome assembly

Hexaploid wheat (17 Gb)
N50 < 10 kb
(IWGSC, Science 2015)

Finger millet (1.5 Gb)
N50 ~ 24 kb
(Hittalmani et al. 2017)

>100X

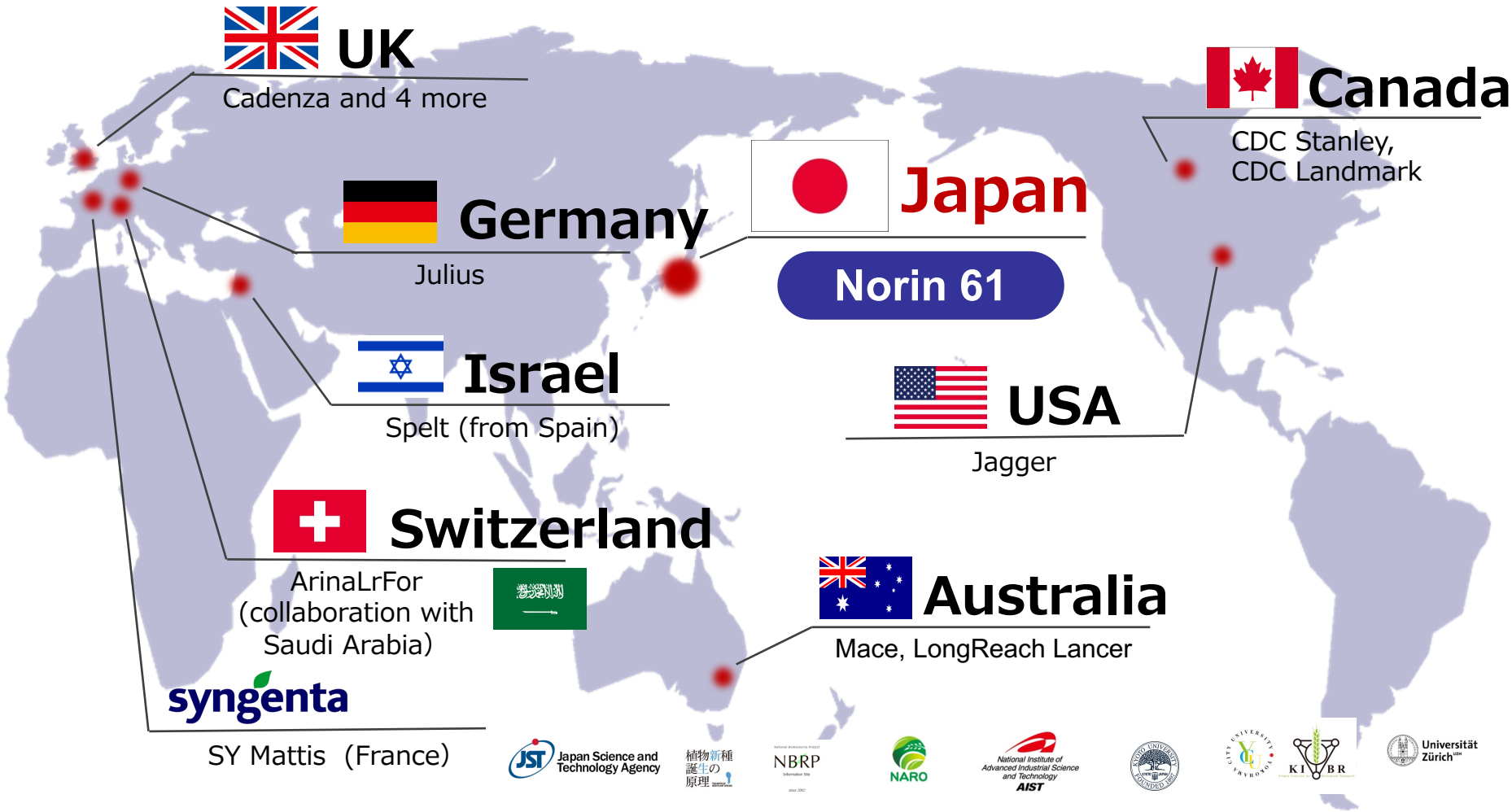
Brassica juncea
N50 ~ 1.5 Mb (Yang et al. 2016)

Tetraploid wheat (12 Gb)
N50 ~ 7 Mb (Avni et al. 2017)

Hexaploid wheat (Chinese Spring)
N50 ~ 22.8 MB (2018)

Finger millet (1.5 Gb) Bionano
N50 ~ 2.6 Mb
Hatakeyama et al. 2017

Wheat 10+ genomes project

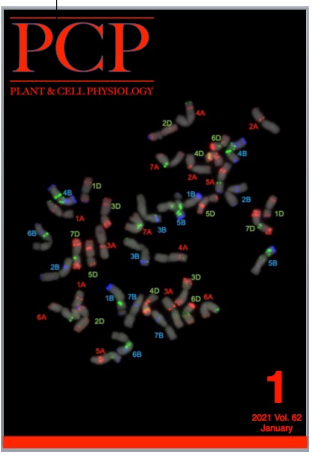


Japanese cultivar as a sole Asian representative

Wheat 10+ genomes project

Walkowiak et al. *Nature* 288: 577, 2020

Shimizu et al. *PCP* online 2020 (draft cover picture)



Article

Multiple wheat genomes reveal global variation in modern breeding

https://doi.org/10.1038/s41586-020-2961-x
Received: 3 April 2020
Accepted: 9 September 2020
Published online: 25 November 2020
Open access
Check for updates

Sean Walkowiak[1,2,41], Liangliang Gao[3,41], Cecile Monat[4,41], Georg Haberer[5], Mulualem T. Kassa[6], Jemima Brinton[7], Ricardo H. Ramirez-Gonzalez[7], Markus C. Kolodziej[8], Emily Delorean[9], Dinushika Thambugala[9], Valentyna Klymiuk[1], Brook Byrns[1], Heidrun Gundlach[5], Venkat Bandi[10], Jorge Nunez Siri[10], Kirby Nilsen[1,11], Catharine Aquino[12], Axel Himmelbach[4], Dario Copetti[13,14], Tomohiro Ban[15], Luca Venturini[16], Michael Bevan[7], Bernardo Clavijo[17], Dal-Hoe Koo[3], Jennifer Ens[1], Krystalee Wiebe[1], Amidou N'Diaye[1], Allen K. Fritz[3], Carl Gutwin[10], Anne Fiebig[4], Christine Fosker[17], Bin Xiao Fu[2], Gonzalo Garcia Accinelli[17], Keith A. Gardner[18], Nick Fradgley[18], Juan Gutierrez-Gonzalez[19], Gwyneth Halstead-Nussloch[13], Masaomi Hatakeyama[12,13], Chu Shin Koh[20], Jasline Deek[21], Alejandro C. Costamagna[22], Pierre Fobert[6], Darren Heavens[17], Hiroyuki Kanamori[23], Kanako Kawaura[15], Fuminori Kobayashi[23], Ksenia Krasileva[17], Tony Kuo[24,25], Neil McKenzie[7], Kazuki Murata[26], Yusuke Nabeka[26], Timothy Paape[13], Sudharsan Padmarasu[4], Lawrence Percival-Alwyn[18], Sateesh Kagale[6], Uwe Scholz[4], Jun Sese[25,27], Philomin Juliana[28], Ravi Singh[28], Rie Shimizu-Inatsugi[13], David Swarbreck[17], James Cockram[18], Hikmet Budak[29], Toshiaki Tameshige[15], Tsuyoshi Tanaka[23], Hiroyuki Tsuji[15], Jonathan Wright[17], Jianzhong Wu[23], Burkhard Steuernagel[7], Ian Small[30], Sylvie Cloutier[31], Gabriel Keeble-Gagnère[32], Gary Muehlbauer[19], Josquin Tibbets[32], Shuhei Nasuda[26], Joanna Melonek[30], Pierre J. Hucl[1], Andrew G. Sharpe[20], Matthew Clark[16], Erik Legg[33], Arvind Bharti[33], Peter Langridge[34], Anthony Hall[17], Cristobal Uauy[7], Martin Mascher[4,35], Simon G. Krattinger[8,36], Hirokazu Handa[23,37], Kentaro K. Shimizu[13,15], Assaf Distelfeld[38], Ken Chalmers[34], Beat Keller[8], Klaus F. X. Mayer[5,39], Jesse Poland[3], Nils Stein[4,40], Curt A. McCartney[9]✉, Manuel Spannagl[5]✉, Thomas Wicker[8]✉ & Curtis J. Pozniak[1]✉

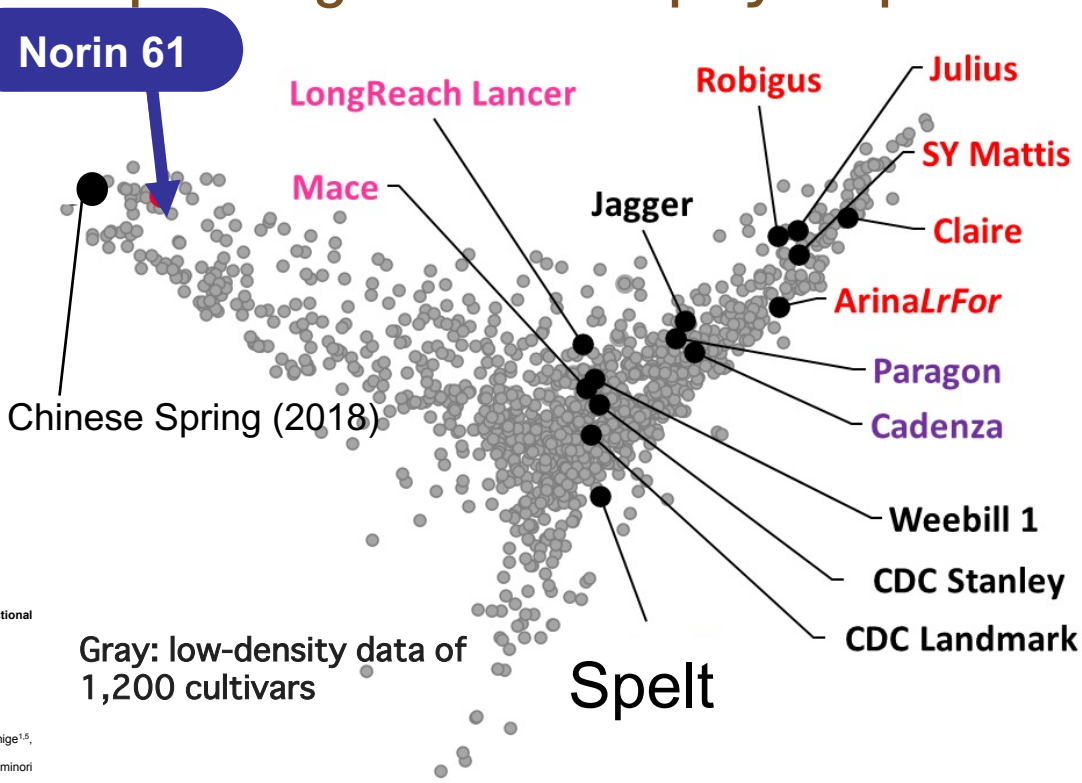
De Novo Genome Assembly of the Japanese Wheat Cultivar Norin 61 Highlights Functional Variation in Flowering Time and *Fusarium* Resistance Genes in East Asian Genotypes

Running head: Genome assembly of the Japanese wheat Norin 61

Authors:

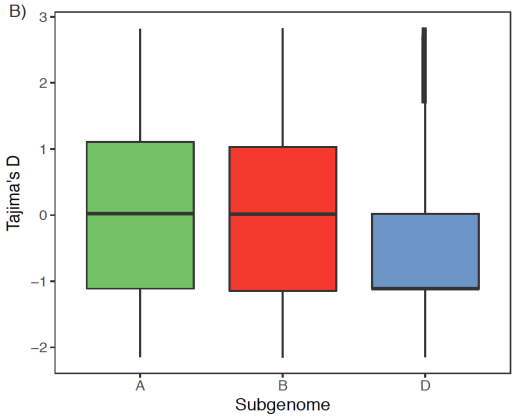
Kentaro K. Shimizu[1,2*#], Dario Copetti[2,3#], Moeko Okada[2#], Thomas Wicker[4], Toshiaki Tameshige[1,5], Masaomi Hatakeyama[2,6], Rie Shimizu-Inatsugi[2], Catharine Aquino[6], Kazusa Nishimura[7], Fuminori Kobayashi[8], Kazuki Murata[9], Tony Kuo[10,11], Emily Delorean[12], Jesse Poland[12], Georg Haberer[13], Manuel Spannagl[13], Klaus F. X. Mayer[13,14], Juan Gutierrez-Gonzalez[15], Gary J. Muehlbauer[15], Cecile Monat[16], Axel Himmelbach[16], Sudharsan Padmarasu[16], Martin Mascher[16], Sean Walkowiak[17,18], Tetsuya Nakazaki[7], Tomohiro Ban[1], Kanako Kawaura[1], Hiroyuki Tsuji[1], Curtis Pozniak[17], Nils Stein[16,19], Jun Sese[9,20*], Shuhei Nasuda[9*], Hirokazu Handa[8,21*]

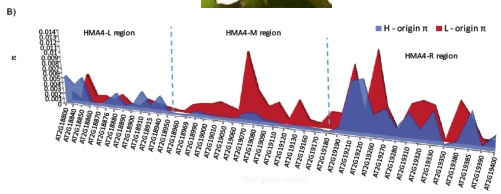
PCA plot of genome wide polymorphisms



Different signature of selection among duplicated genes

Tajima's *D* (signature of selection)





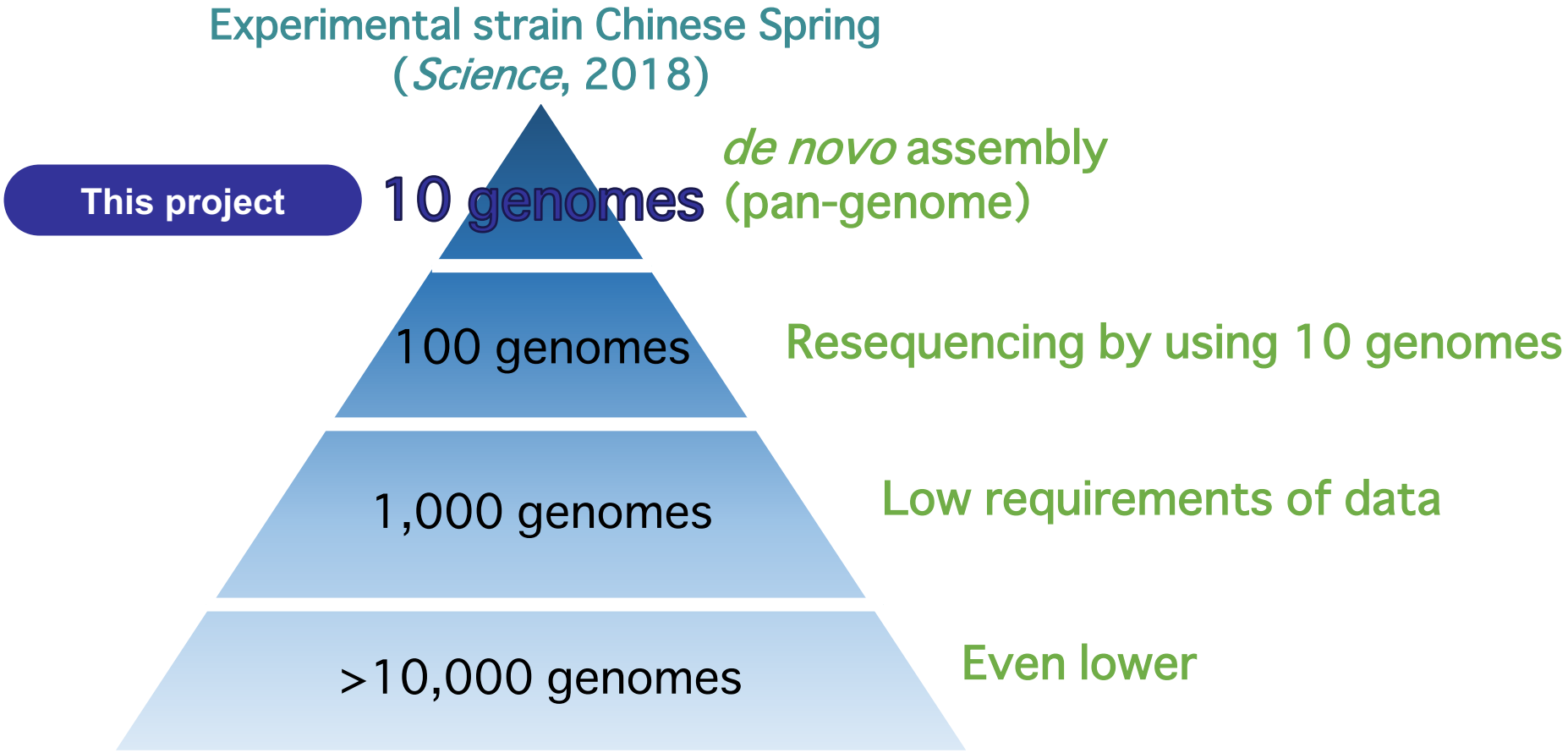
Paape et al. *Nature Comm* 9: 3909, 2018

Low correlation between duplicated copies	
Crop (bread wheat)	Natural (*Arabidopsis kamchatica*)
r = 0.02-0.06	r = 0.05

Walkowiak et al. *Nature* 288: 577, 2020

The same pattern in natural and crop polyploid species

Era of multiple genomes



Basic and applied researches