Provenance trial networks as a tool for biochemical and molecular genetics of forest trees

Berthold Heinze

BFW – TBX P02 – Vienna, Austria

I - Field trials as a “quick and easy” way to collect material

- collect diverse material for genetic marker studies in one place

pros:

- many diverse sources at one place
- replicated (other labs can use the same material) - standardisation & comparison
- relevant for practical purposes – hope to distinguish better and worse provenances with markers

Lagercrantz and Ryman 1988, 1990

- first to assess range-wide variation in a forest tree with isoenzyme (allozyme) markers
  - Norway spruce IUFRO 1964/68 trial in Sweden
- key innovation: using diploid material from buds for analysis
- multivariate trends in accordance with geography

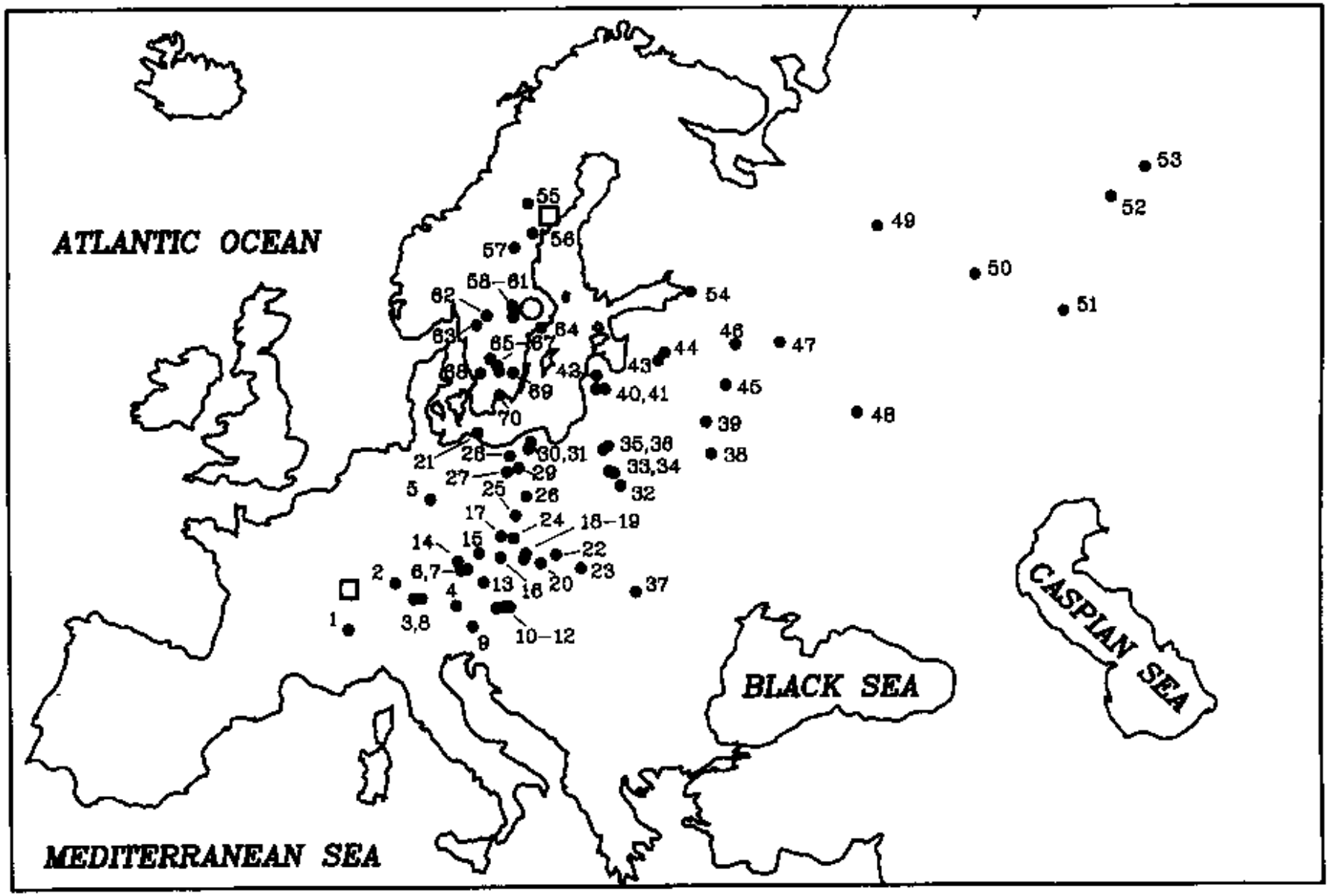


FIG. 1. Geographic locations of the original seed collections of Norway spruce (see Table 1). The open squares indicate experimental sites for morphological data from the field trials of the IUFRO 1964/68 program; the open circle indicates the experimental site for the nursery experiment in central Sweden.

Further examples

- Prus-Glowacki and Bernard 1984,
- Oleksyn et al. 1994 *(Pinus sylvestris):*
  - correlation of genetic data with pollution of the field trial site
- Kannenberg and Gross 1999 *(Picea abies):*
  - geograpic patterns at some loci
  - higher variation in the North and in the Balkans
- Mihai and Teodosiu 2009 (*Larix decidua*):
  - high diversity at the edge of the range

Kannenberg and Gross 1998

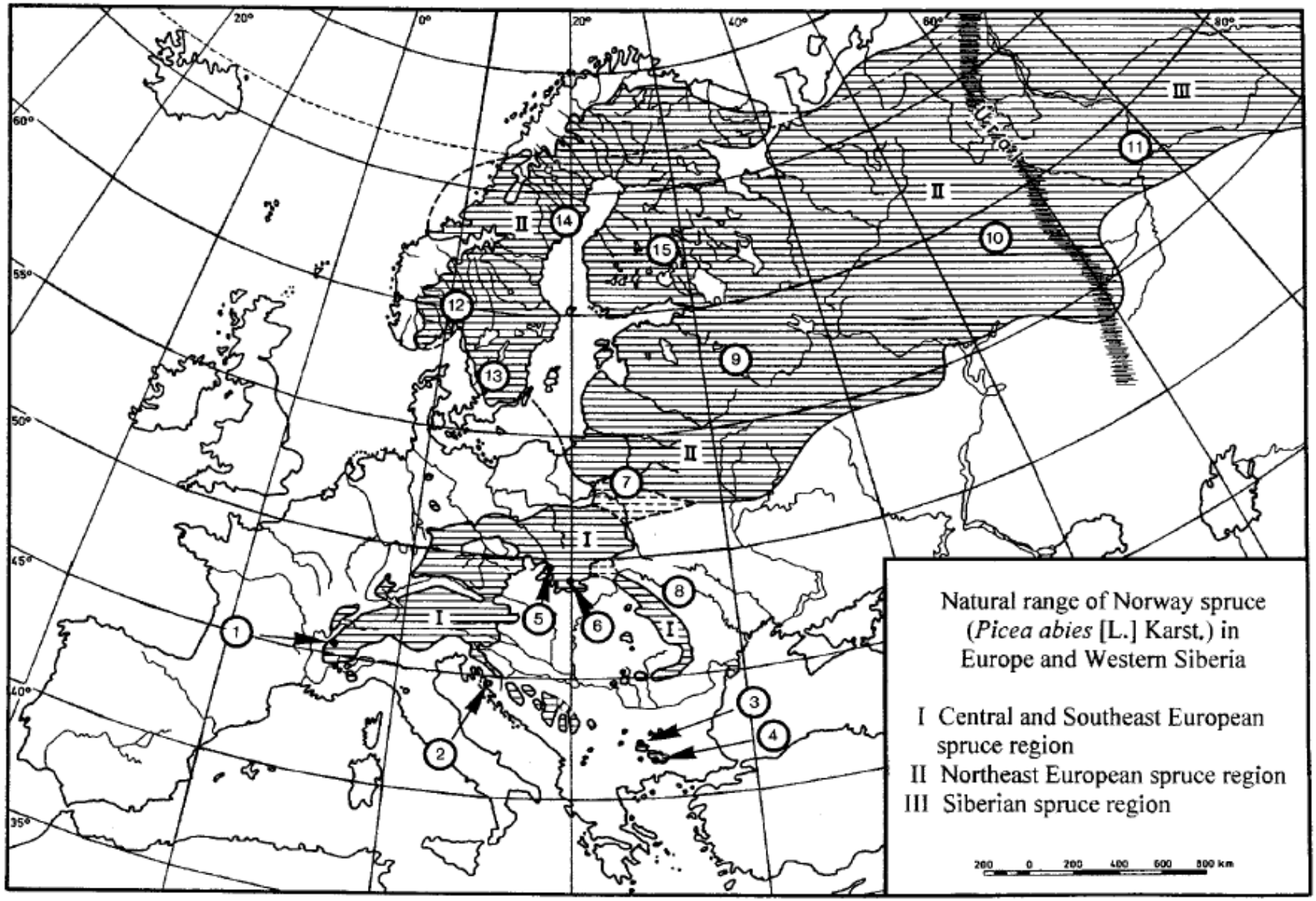


Figure 1. – The European natural range of Norway spruce [from SCHMIDT-VOGT (1977), slightly modified] and the places of origin of the 15 spruce provenances investigated.

Nice example from Poland

- Chalupka et al. 2008 (*Picea abies*):
- reconstitution of Kolonowskie seed source
- original stand of IUFRO seed collection disappeared
- source was very good at many test sites
- seed orchard constructed from offspring genotypes in tests
- confirmed with genetic markers

Other types of markers in traditional studies

- marker type is largely irrelevant from the point of view of trial management
- other nuclear DNA markers:
  - Perry et al. 1999, *Picea abies*
  - sequence-tagged sites (PCR [RFLP])
- chloroplast microsatellites:
  - Vendramin et al. 2000, *Picea abies*
  - geographic variation in congruence with only two glacial refugia
- mitochondrial minisatellites:
  - Sperisen et al 2001, *Picea abies*
  - confirmed two glacial refugial populations colonizing Europe

Further example

- chloroplast and mitochondrial markers combined:
  - Gugger et al. 2010, *Pseudotsuga menziesii*
  - differentiation of Rocky Mountain populations, but not those at the coast
  - zone of introgression / hybridization
  - use this information to trace origins of early introductions in Europe?

B F W

Department of Genetics

Disadvantages

- exact identification of source
  - especially in older trials
  - area/region vs. stand
- exact descent of material
  - how many mothers - which is which?
- source material may have disappeared
  - seed stands cut for timber
- possible natural genetic selection in the nursery /at the trial site
- comprehensiveness (range-wide?)

Disadvantages - examples

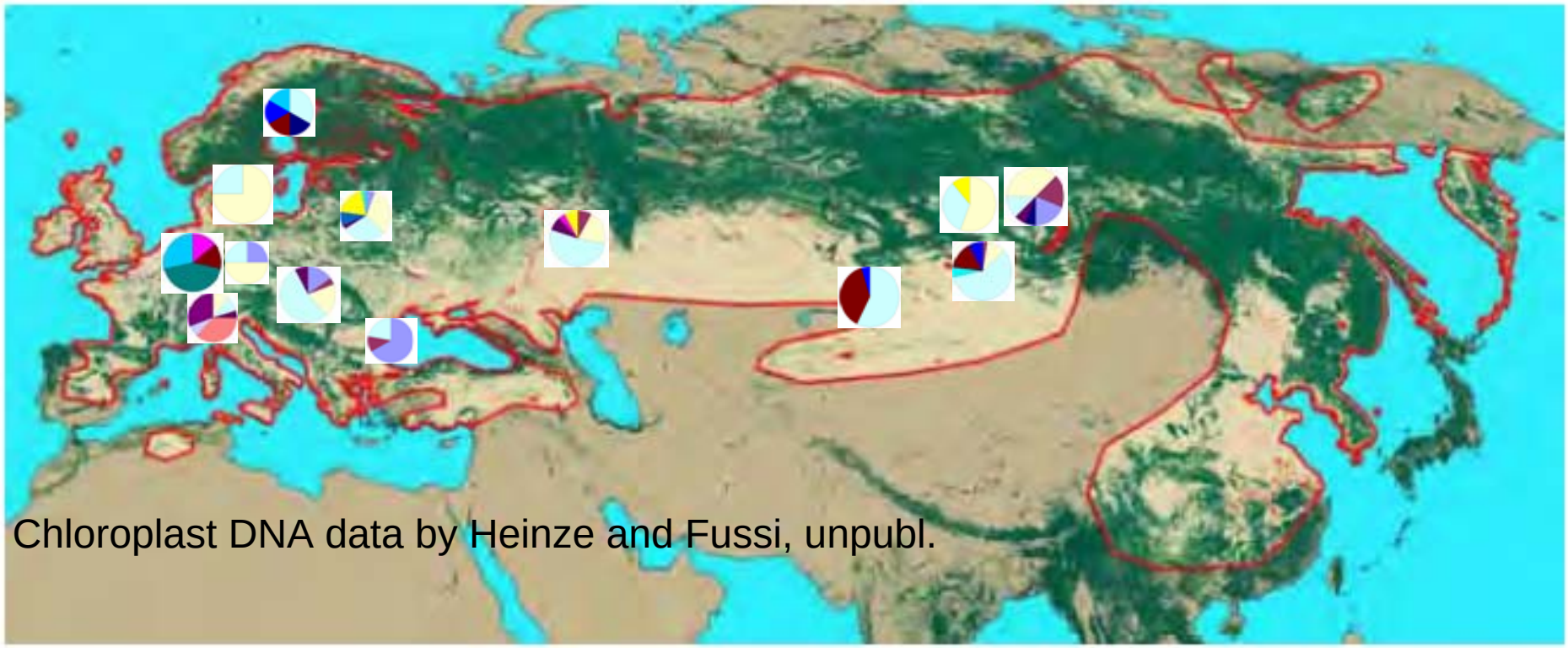
- Cieslar 1905 *Quercus robur*
  - (Cieslar 1923)
  - 1 or 2 mother trees only
  - no repetitions
- pre-IUFRO trials in general
  - often inferior statistical design
  - sources not traceable any more?
- IUFRO trial series restricted to few species
  - spruce, larch, Doug fir
- RAP Fraxinus – not range-wide

Alternatives for obtaining diverse material

- request seeds (or collect yourself)
  - preferred for conifers
  - haploid megagametophytes
- visit stands
  - preferred for controlling relatedness of material
  - e.g. 30/50 m between sampled trees
- correspondents
  - dried leaf material in a letter
  - leaves in silica gel

Example: *Populus tremula* range

- would be impossible to visit multiple sites
- nor to send seeds easily



Map 18. Natural range of *Populus tremula* in Eurasia and Africa. Redrawn from Fenaroli and Gambi (1976).

„Added value“ of large trial network?

- not really present yet
- multiple-site studies are rare
- multiple-lab studies are rare
- has the value not yet been realised?
  - selection effects at different sites?
  - pedigree reconstruction?
  - genetic diversity and plasticity?

Selection, adaptation and epigenetic effects

- seedlings planted in various climates may undergo selection
- difficult to disentangle selection and local adaptation effects
  - first vs. further generations?
- epigeneitc effects described in *Picea abies*
  - T. Skroppa, O. Johnson et al.
  - seedlings behave different if harvested in different climate, but from identical trees
  - Hungarian example – Ujvari Jarmay and Ujvari 2006:
  - *Picea abies* seeds harvested in IUFRO trial
  - selected mother trees often exceeded growth of local material
  - well-known „maternal effect“ (seed nutrition after-effects)
  - evident in high altitude *Picea abies* in the Alps

Little "marking" capacity for really interesting growth traits

- incongruence between observable growth and marker patterns (in some examples)
- often low Fst vs. high Qst
  - little genetic differentiation,
  - high quantitative variation
- reasons?
  - too few markers
  - selectively neutral markers
  - too simple models of inheritance
    - polygenic traits
    - more complex genetic interactions

II - The dawn of the age of genomics

http://www.mansfield.ohio-state.edu/~sabedon/2001_dawn05.jpg

II - The dawn of the age of genomics

- genetic mapping
  - required family pedigrees, not provenances
- maps of markers only, initially
- then QTLs:
  - quantitative trait loci
  - chromosome regions with statistical correlation to quantitatively measured traits
- progeny trials more interesting

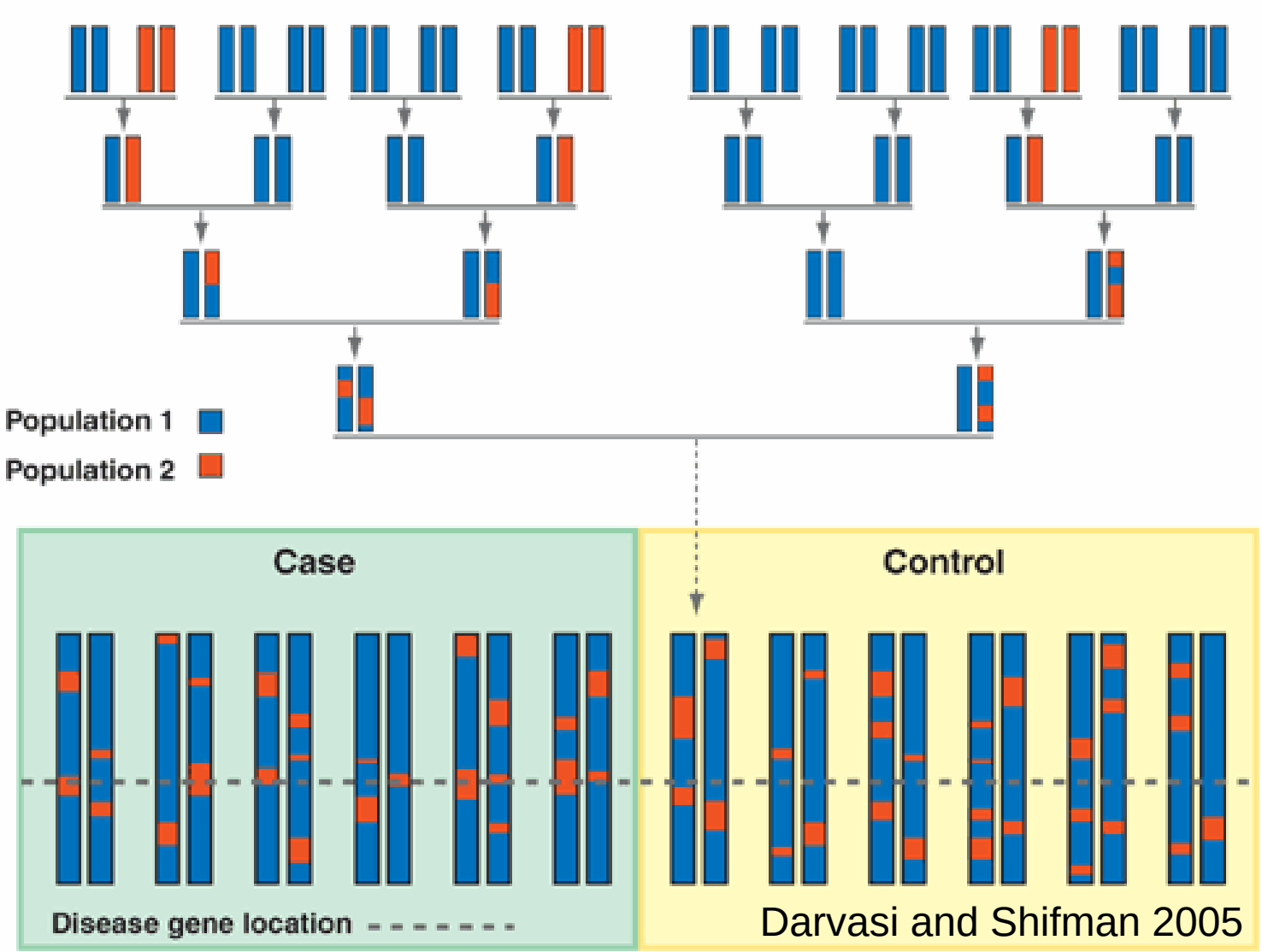
Problems with QTL mapping

- transferability:
  - markers or traits or QTLs (or all of those) not always transferable from one family to the next
  - from one experiment to the next one
  - effect of deleterious alleles in some families
    - vs. real superior alleles
  - interactions (genetic epistasis) are broken in a new genetic background)

Alternatives from human genetics

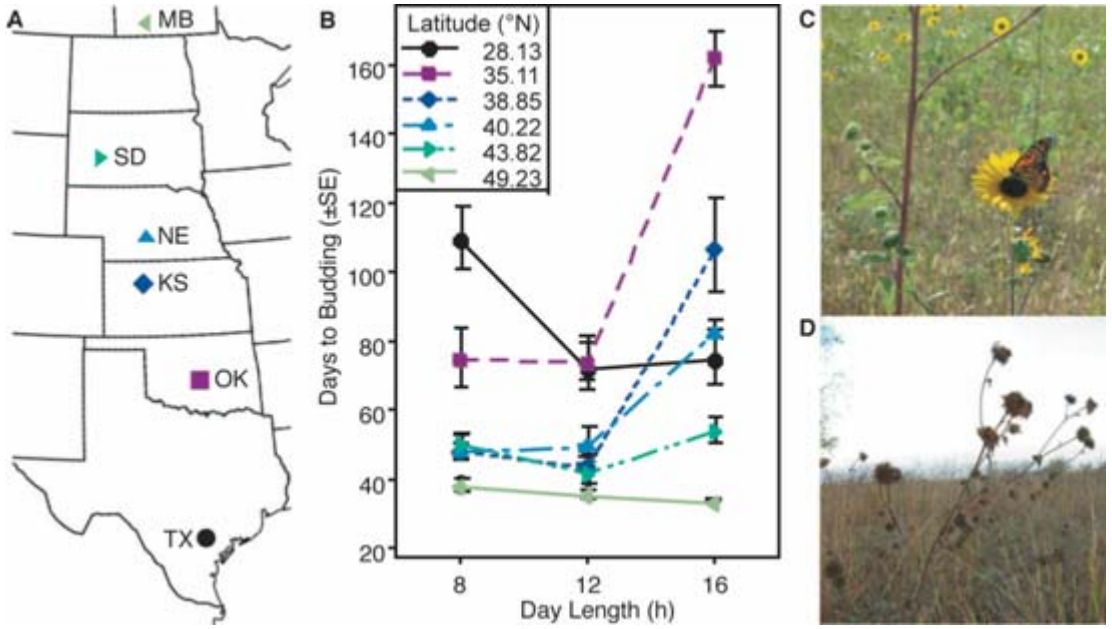
- building large pedigrees is also not feasible
- admixture mapping:
- linkage disequilibrium building up through natural hybridization and backcrossing

Alternatives from human genetics



Examples in plants – Loren Rieseberg‘s lab

- work in hybrid sunflower
- backcrosses loose most genes from other species
- but retain the ones that give them an advantage



http://www3.botany.ubc.ca/rieseberglab/research.html

III - Another alternative: association studies

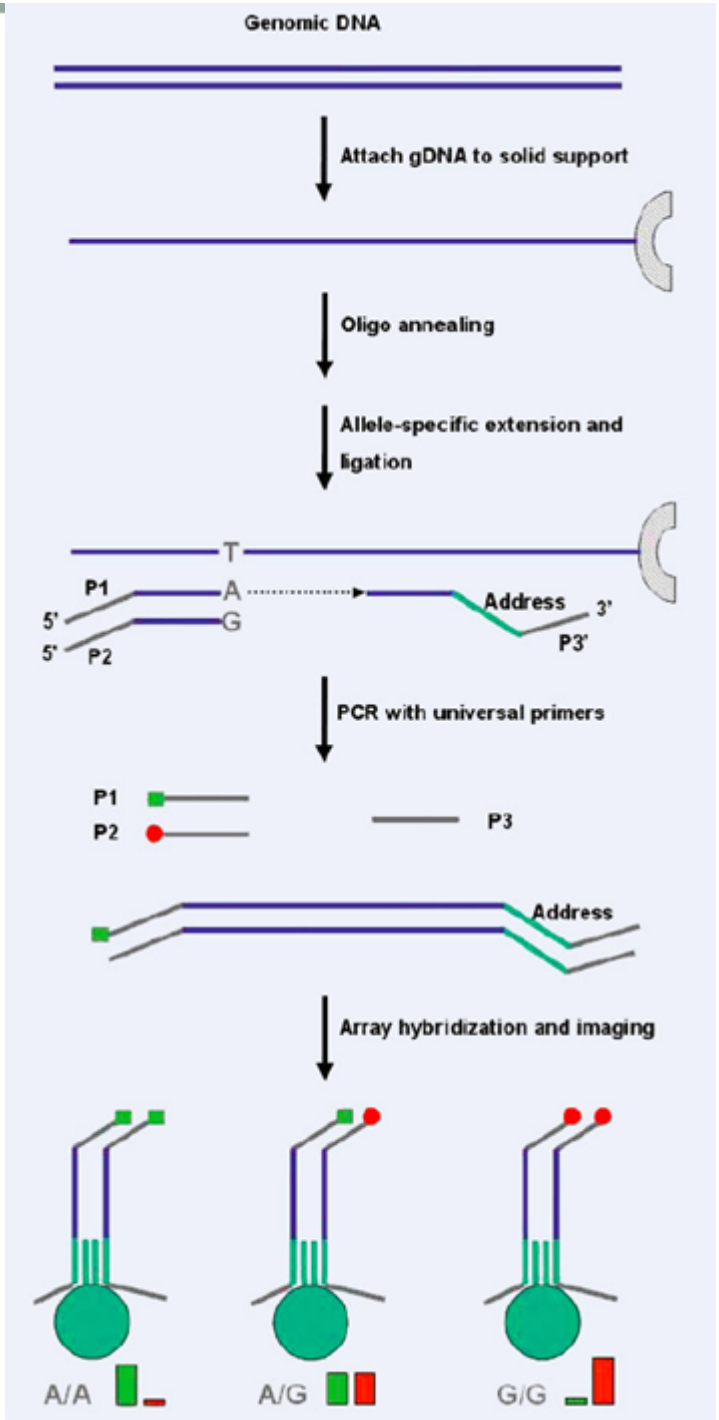
- simple correlations between markers and traits
- going back to the original idea of genetic markers
- at candidate genes
- across the whole genome
  - *Arabidopsis* and other models
- simple, but what are the problems?

Digression - technical advances

- next generation sequencing
  - new sequencing methods for very high throughput
- massively parallel SNP assays
  - assess hundreds of single nucleotide polymorphisms in hundreds of samples
- methods often available from larger centres or specialised companies

Illumina Golden Gate assay

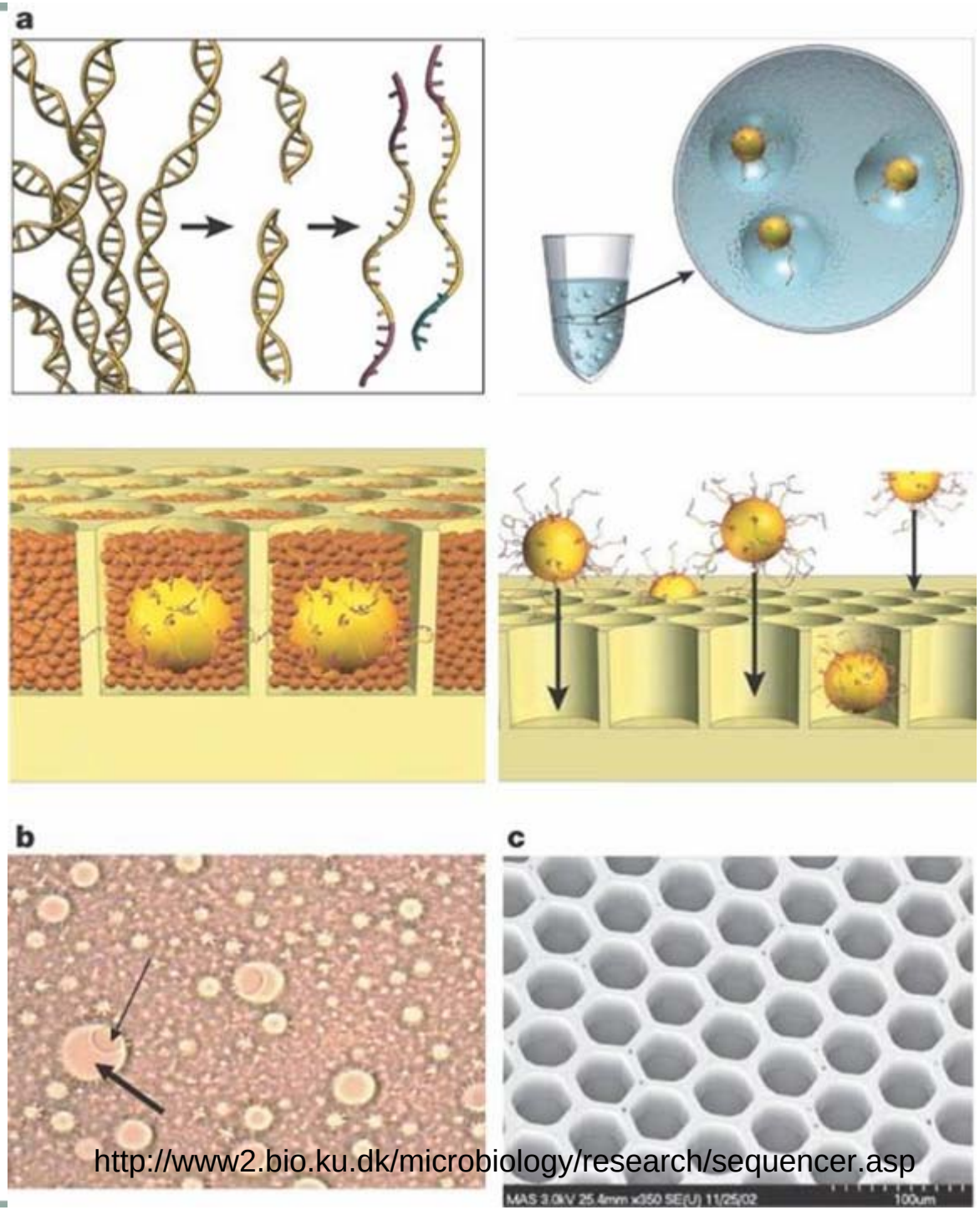
- 1536 pre-defined SNPs in one run
- hundreds (thousands) of individuals



http://www.genomecenter.ucdavis.edu/dna_technologies/illumina.html

Next generation sequencing technology example

- Roche/454 pyro-sequencing



http://www2.bio.ku.dk/microbiology/research/sequencer.asp

Table 1 | **Comparison of next-generation sequencing platforms**

Platform	Library/ template preparation	NGS chemistry	Read length (bases)	Run time (days)	Gb per run	Machine cost (US$)	Pros	Cons	Biological applications	Refs
Roche/454's GS FLX Titanium	Frag, MP/ emPCR	PS	330*	0.35	0.45	500,000	Longer reads improve mapping in repetitive regions; fast run times	High reagent cost; high error rates in homo-polymer repeats	Bacterial and insect genome *de novo* assemblies; medium scale (<3 Mb) exome capture; 16S in metagenomics	D. Muzny, pers. comm.
Illumina/ Solexa's GA_{II}	Frag, MP/ solid-phase	RTs	75 or 100	4‡, 9§	18‡, 35§	540,000	Currently the most widely used platform in the field	Low multiplexing capability of samples	Variant discovery by whole-genome resequencing or whole-exome capture; gene discovery in metagenomics	D. Muzny, pers. comm.
Life/APG's SOLiD 3	Frag, MP/ emPCR	Cleavable probe SBL	50	7‡, 14§	30‡, 50§	595,000	Two-base encoding provides inherent error correction	Long run times	Variant discovery by whole-genome resequencing or whole-exome capture; gene discovery in metagenomics	D. Muzny, pers. comm.
Polonator G.007	MP only/ emPCR	Non-cleavable probe SBL	26	5§	12§	170,000	Least expensive platform; open source to adapt alternative NGS chemistries	Users are required to maintain and quality control reagents; shortest NGS read lengths	Bacterial genome resequencing for variant discovery	J. Edwards, pers. comm.
Helicos BioSciences HeliScope	Frag, MP/ single molecule	RTs	32*	8‡	37‡	999,000	Non-bias representation of templates for genome and seq-based applications	High error rates compared with other reversible terminator chemistries	Seq-based methods	91
Pacific Biosciences (target release: 2010)	Frag only/ single molecule	Real-time	964*	N/A	N/A	N/A	Has the greatest potential for reads exceeding 1 kb	Highest error rates compared with other NGS chemistries	Full-length transcriptome sequencing; complements other resequencing efforts in discovering large structural variants and haplotype blocks	S. Turner, pers. comm.

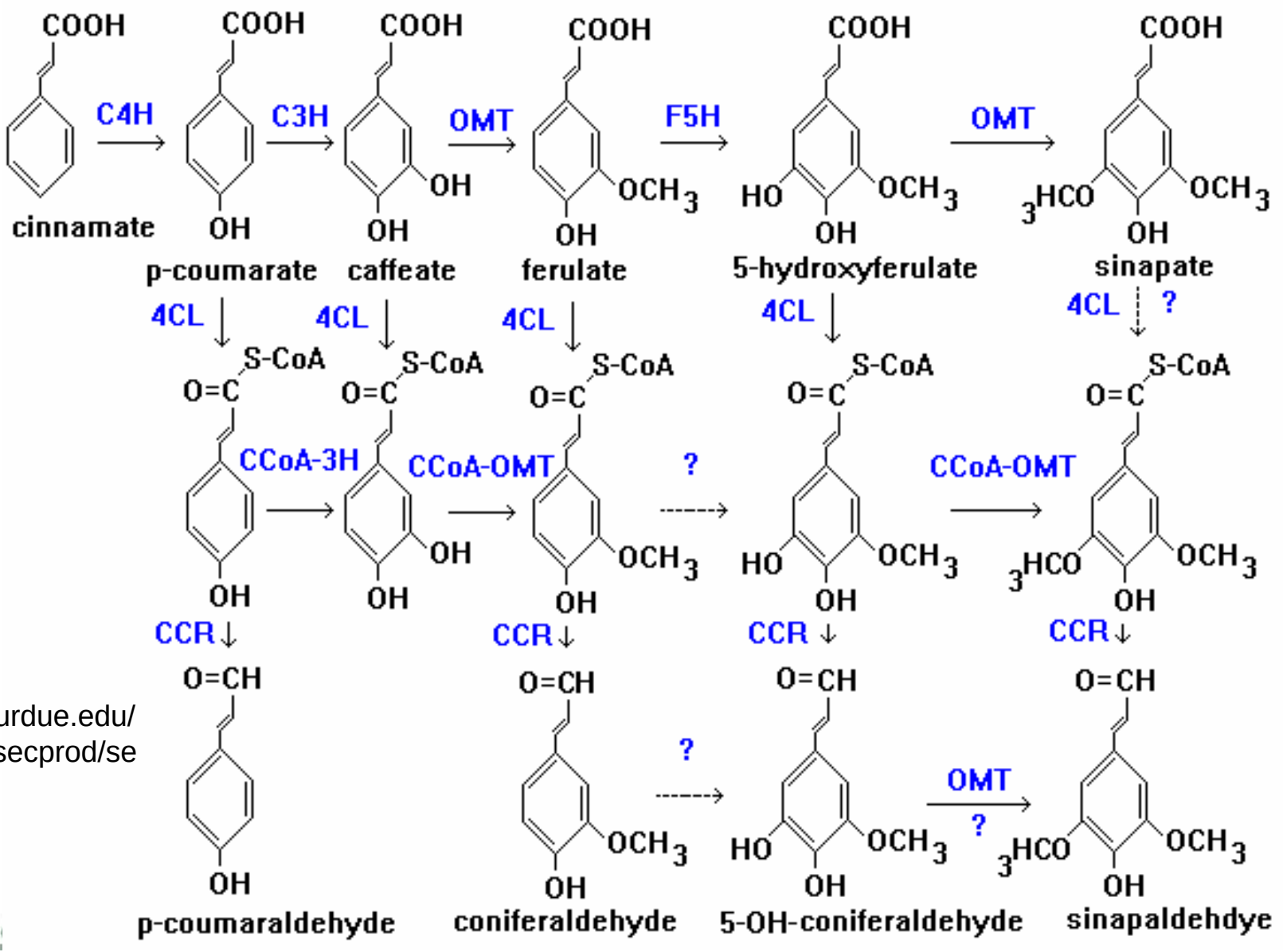
*Average read-lengths. ‡Fragment run. §Mate-pair run. Frag, fragment; GA, Genome Analyzer; GS, Genome Sequencer; MP, mate-pair; N/A, not available; NGS, next-generation sequencing; PS, pyrosequencing; RT, reversible terminator; SBL, sequencing by ligation; SOLiD, support oligonucleotide ligation detection.

Metzker 2010

How to do an association study

- collect material
- measure phenotypes
  - height, diameter, diseases, ...
- analyse as many markers as possible:
- candidate genes
  - for biological function
  - gene expression
  - from model organisms
  - from QTL regions
  - ...
- alternatively – whole genome sequencing
  - individual genomes in *Arabidopsis*
  - pools for other organisms (Futschik and Schlötterer 2010 in press)

Example - lignin pathway genes



http://www.hort.purdue.edu/rhodcv/hort640c/secprod/se00016.htm

Genetic analysis in association studies

- mostly done by sequencing genes
  - PCR & sequencing
- or analysis of SNPs
  - sometimes a selection only
- next generation sequencing for sequence / SNP discovery
  - but not yet for re-sequencing = analysing the individual samples

How to do an association study (II)

- assess structure in the sample
- need to control for population substructure / family structure
  - e.g. STRUCTURE, pedigree reconstruction
- calculate statistical associations
  - dedicated software
  - special tests if structure is present
- verify in independent sample
  - e.g, 2/3 of sample in association
  - and 1/3 of sample for verification

Advantages of association studies

- ease of the approach for sampling
- inherently simple approach
- no building of pedigrees necessary
  - but family pedigrees can enhance the study

IV - Examples of association studies in trees (overview)

Demography

Heuertz et al. 2006
Picea abies
22 loci
excess of rare and high-freq. mutations; bottleneck

Pyhäjärvi et al. 2007, Palmé et al. 2008
Pinus sylvestris
16 candidate genes / EST databases
demography / selective sweeps

Eveno et al. 2007
Pinus pinaster
11 candidate genes
„outlier“ loci

Keller et al. 2010
Populus balsamifera
412 SNPs in 474 individuals + 11 sequenced genes in 94 individuals
3 geographical clusters; massive expansion inferred (after Ice Age)

Local adaptation

Ingvarsson et al. 2008, Luquez et al. 2008, ...

Populus tremula

77 gene fragments

excess of low-frequency mutations; bottleneck; association of flowering pathway genes with bud set (PHYB)

Namroud et al. 2008

Picea glauca

534 SNPs in 345 expressed genes

genes involved in local adaptation of some populations (e.g. drought, heat)

Holliday 2009 (dissertation)

Picea sitchensis

candidate genes from microarray studies; 768 SNPs

widespread purifying selection; some positive / diversifying selection; 28 associations for cold hardiness and budset (explained ~ 30% of phenotypic variation in mapping population from 12 geographical locations)

Wood traits

Gonzalez-Martinez et al. 2007, 2008; Eckert et al. 2010 in press

Pinus taeda

SNPs in up to 3059 genes

wood properties; carbon isotype discrimination; abiotic stress response; expansion from Mexico and Florida

Eckert et al. 2009a, b

Pseudotsuga menziesii

384 SNPs in 117 candidate genes / 121 candidate genes

cold-hardiness traits – 30 associations in 12 genes; 7 markers differentiated coast / interior; small effects of genes; selective sweeps at 3-8 loci; bottleneck

Dillon et al. 2010 in press

Pinus radiata

149 SNPs in cell wall candidate genes

10 significant associations with wood property traits

Characteristics of first generation of studies

- using traditional Sanger sequencing of some candidate genes and / or
- SNP detection panel
  - only a handful of samples
- followed by SNP assay on many individuals
- testing for deviation from neutrality
  - genes or alleles that show reduced or enhanced diversity
  - „footprints of selection“
  - „selective sweeps“
- testing for association with „geography“, wood traits

Issues with association studies

- sequences/primers not available for all species
- when testing many markers in many individuals, how to distinguish false positives from true association?
- association (statistical correlation) does not mean causal explanation
- often only a low percentage of variation explained by the markers/alleles/ SNPs
  - few percent, even if added
- would make marker-based selection inefficient

Recent exception - Pär Ingvarsson - *P. tremula*

- when considering also LD between markers, they explain up to 50% of phenotypic variation !
- approach suggested by Lewontin and Krakauer, 1970ies
  - P. Ingvarsson, @ EVOLTREE conference El Escorial, Spain, June 2010

HK Wildermieming/T (400-900 m), 22-jährig

Plantage Hamet (P3) - Lammerau (400-700 m), 22-jährig

Conclusions

- genome-wide („genomic“) studies will hopefully reveal genetic control of traits in many species soon
- technology advances make it possible to study many genes / whole genomes
- experimental networks are an ideal basis for such studies
- both provenance and progeny trials can be used
  - mix of unrelated material and crosses for plants
  - Myles et al. 2009
- basic research into gene function is necessary before gene markers can be used for selection

Phenotyping
(measuring, observing, assessing, testing, counting ...)
= „phenomics“
will become more and more important for genetic studies as genotyping becomes easier

Some of the studies are based on pedigrees, but ...

... does this mark the return of the provenance trials?

The return of the son of the provenance trial: genetic association studies in trees

Acknowledgement

- Wladislaw and Jan for the invitation ...
- ... and inspiration!
- Michael Mengl for literature hints
- Lambert Weißenbacher for provenance trial pictures
- Christian Lexer and Barbara Fussi for collaboration in *P. tremula / P. alba*

- thank you for your kind and lasting attention