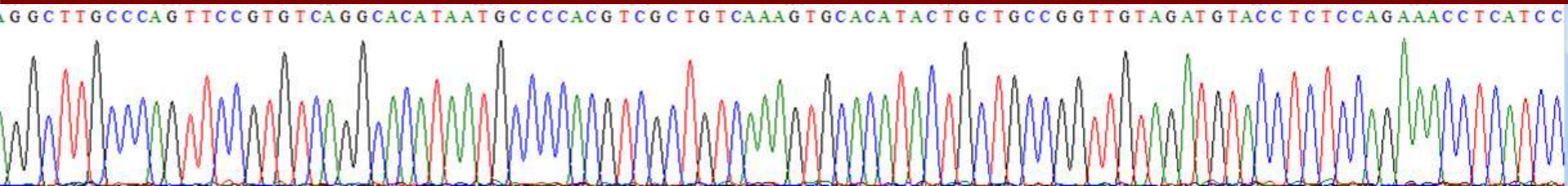
Molecular DNA markers suitable for population/stock delimitation and assignment

Ralph Tiedemann, University of Potsdam



Concept of molecular markers

Basic idea:

- Individuals are genotyped at particular genetic loci.
- State/variation/divergence at these loci is interpreted as to representing state/variation/divergence of entire population(s).

Assumptions:

- Gene flow among demographically independent units (stocks/populations) is less than expected at random.
- Hence, random mutations and genetic drift will translate into measurable genotype frequency differences among populations

Molecular markers in population assessment

Biological marker characteristics

- Function of marker locus (if any)

  Potential influence of natural selection

- Marker inheratince

  clonal (uniparental) or Mendelian

- Evolutionary mode

  Length variation (indels)/ point mutations

- Evolutionary rate

  mutation/substitution rate

Molecular markers in population assessment

Technical marker characteristics

- Accessibility

  Is the marker established for the focus taxon ?

- Avoidance of ascertainment bias

  Bias-free with regard to the populations of interest ?

- Reliability

  Is suitable organismal material available ?

  Is the genotype reliably assessed (e.g., no null alleles)?

- Repeatability/Transferability

  Are data comparable across experiments/labs ?

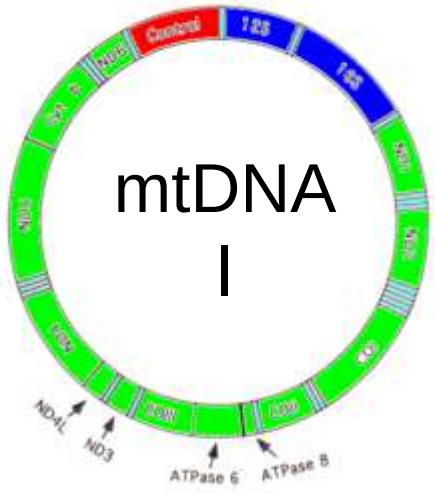
- Costs

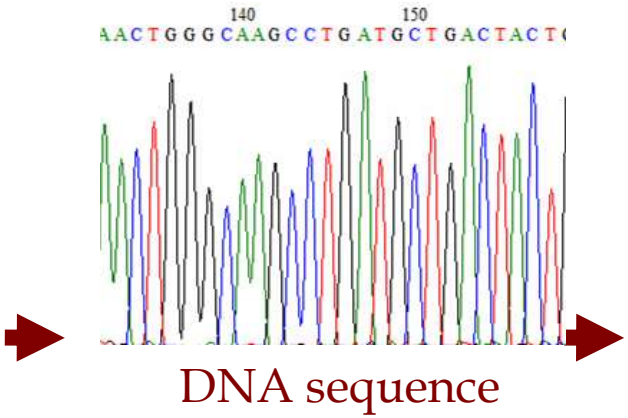
  Is the approach affordable ?

Molecular markers in evolutionary research

Level	mtDNA-sequence	ncDNA-sequence	Multilocus-Fingerprints	Singlelocus-Fingerprint (microsats)	RAPD AFLP	SNPs
Identity	(+)	–	+	++	+	+
Paternity/kinship	(+)	–	+	++	+	+
Populations	++	(–)	(+)	++	+	++
Species	++	+	–	(+)	–	(+)
Genera	++	+	–	(+)	–	–
Families/orders	+	++	–	–	–	–
Classes/Phyla	(+)	++	–	–	–	–

Sequence analysis of mitochondrial DNA:





DNA sequence

BAL 1	CTATCTAGCTTCCACCTACTTA
BAL 2	T.....C..T.....C.
BAL 3	T.....
BAL 4	.....TC..T.....C.
BAL 5	.....C..T.....
BAL 6	T.....G.....
BAL 7	T.....C.TT.....C.
BAL 8	.....C..T.....C.
BAL 9	T.....T.....

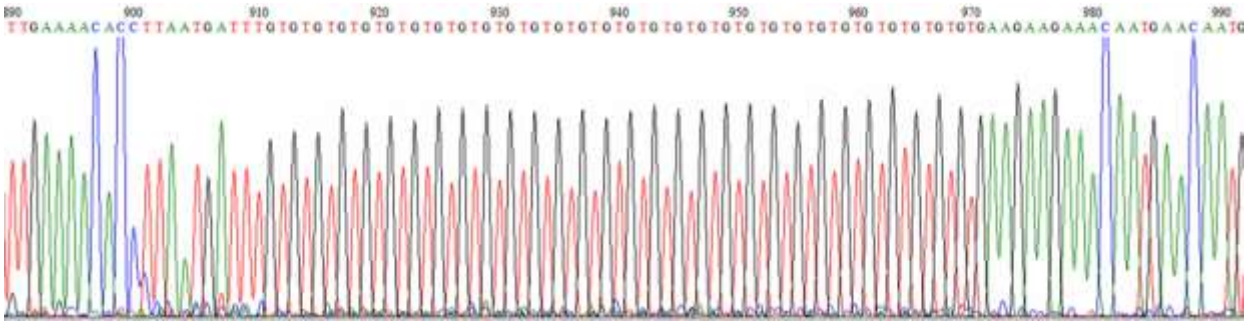
haplotypes

Advantages:
- established for many taxa
- high repeatability across labs
- little requirements on sample quality

Caveats:
- maternal clonal inheritance
- only one locus
- too little variation in some species

Single locus fingerprints (microsatellites):

- Short sequence repeats
- highly variable due to polymerase„slippage“



$(GT)_8$

GTCCGTGTGTGTGTGTGTGTAGCTGT

GTCCGTGTGTGTGTGTGTAGCTGT

$(GT)_7$

Advantages:

- highly polymorphic
- Mendelian inheritance (if autosomal)
- Selectively neutral
- Equal contribution of sexes (if autosomal)
- medium requirements on sample quality

Caveats:

- not directly comparable across labs
- possible mistyping due to stuttering
- possible null alleles

Null alleles:

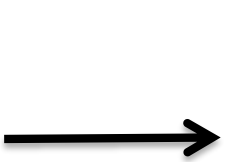
Heterozygous:

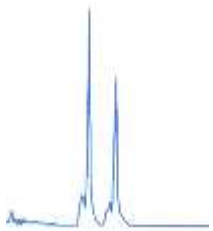
$(GT)_8$

GTCCGT GTGTGTGTGTGTGTGT AGCTGT

GTCCGT GTGTGTGTGTGTGT AGCTGT

$(GT)_7$





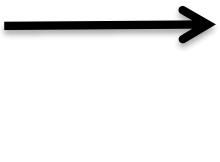
Homozygous:

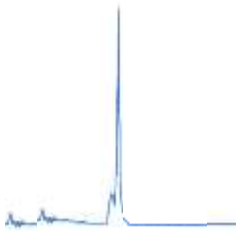
$(GT)_8$

GTCCGT GTGTGTGTGTGTGT AGCTGT

GTCCGT GTGTGTGTGTGTGT AGCTGT

$(GT)_7$





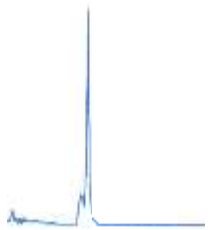
Heterozygous (with null allele):

$(GT)_8$

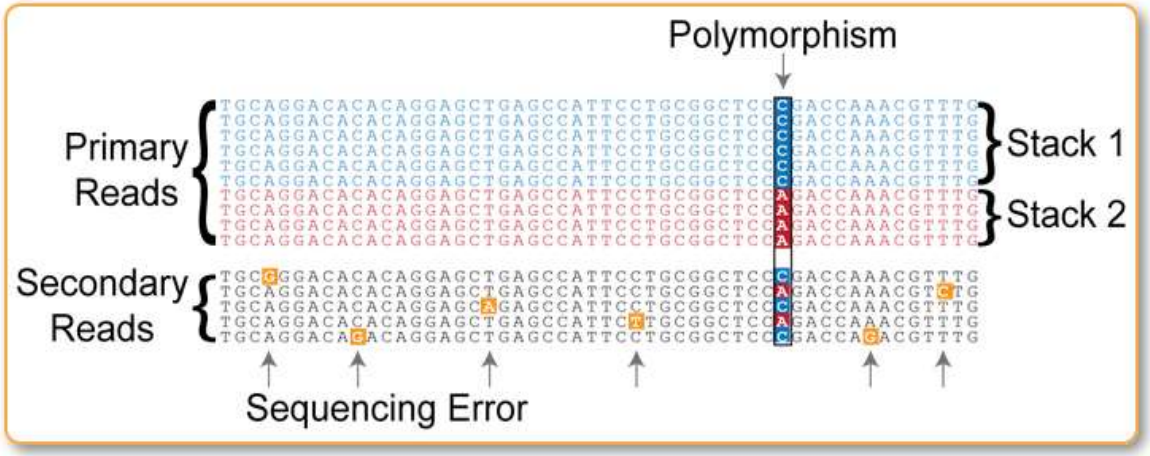
GTTCGT GTGTGTGTGTGTGTGT AGCTGT

GTCCGT GTGTGTGTGTGTGT AGCTGT

$(GT)_7$



Single nucleotide polymorphisms (SNPs):



Sample	Total
Raw reads	303,307,578
Raw read pairs	151,653,789
Average read pairs/sample	3,020,035
Average number loci/sample	374,787
Average number SNPs/sample	40,652
Shared SNPs across samples	1,801

Advantages:

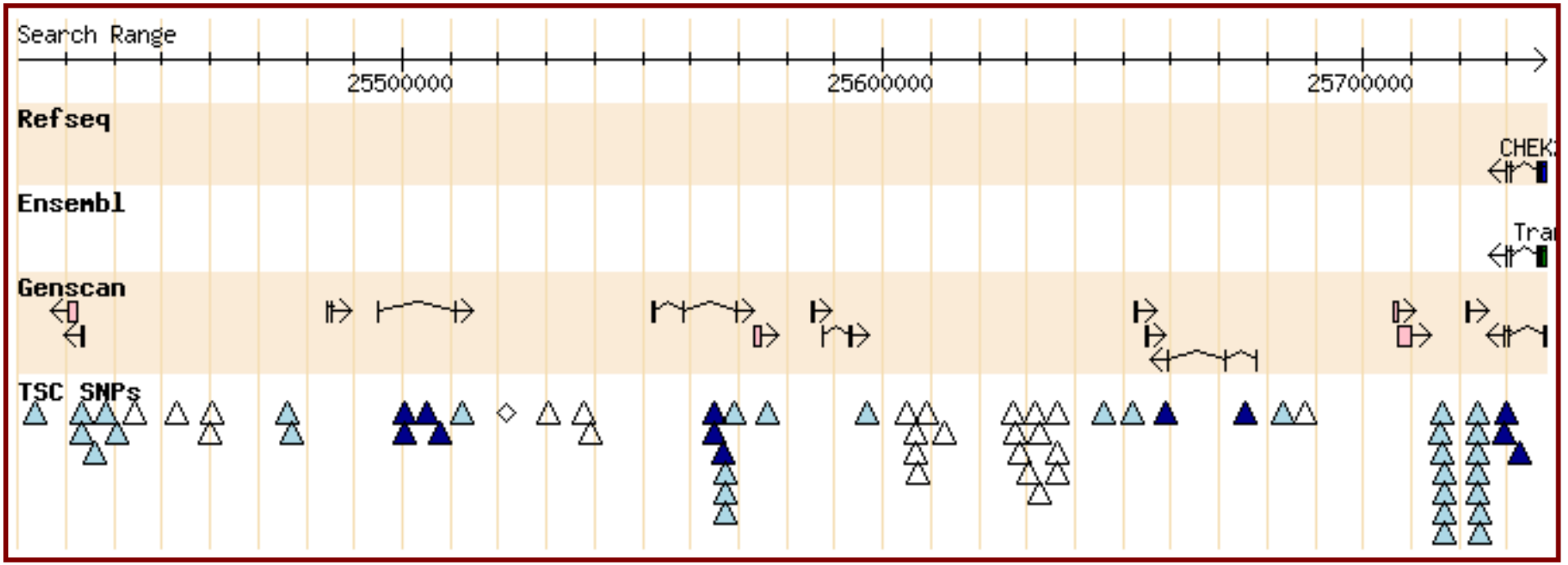
- Mendelian inheritance (if autosomal)
- Equal contribution of sexes (if autosomal)
- high repeatability across labs

Caveats:

- higher requirements on sample quality
- little information per locus
- possible mistyping due to null alleles
- not directly available for many taxa
- potential linkage disequilibrium among loci

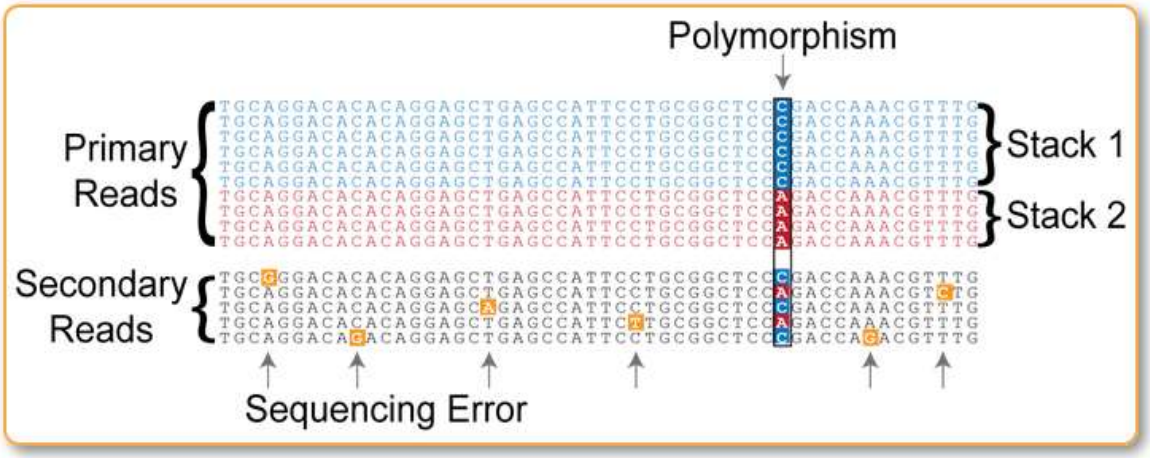
49 samples
100 bp Paired End reads
CUTTERS: MspI, PstI

Mapped SNPs:



SNPs on human chromosome 22 from 25.420 Mb to 25.738 Mb, 317,751 bp window (http://snp.cshl.org)

SNP PoolSeq:



Advantages:

- Cost effective

Caveats:

- high requirements on sample quality
- biallelic inheritance cannot be seen
- no individual genotypes
- potentially differential contribution of individuals
- possible mistyping due to null alleles
- Difficult to tell apart polymerase/sequencing errors and rare alleles

Quality issues in PCR-based genotyping:

Contamination:

- Cross-sample contamination
- Contamination with PCR-products

Mispriming:

- Amplification of „wrong“ loci
- Non-amplification due to mutations in primer sites („null alleles“)

Genotype errors:

- Polymerase error rate 0,001 - 0,0001
- Overall error rate typically higher (should be estimated

Quality issues in molecular markers:

Markers ideally used for those species/stocks they have been developed for

If not:

- Potentially biased assessment of genetic variation/stock structure (ascertainment bias)
- Increased likelihood of amplification of „wrong“ loci (mispriming)
- Increased likelihood of mutations in primer sites (null alleles)

Molecular markers in population assessment

Microsatellites/SNPs (and sometimes mtDNA) are suitable markers for population/stock assignment, if

- Markers have been validated for the stocks of interest

  No ascertainment bias, no null alleles
- Marker inheritance is taken into account

  clonal (uniparental) or Mendelian
- A sufficiently large number of loci/informative SNPs is screened (ideally in individual specimens)
- Linkage disequilibrium is taken into account
- General rules of data quality are followed