ProCoGen

An introduction to conifer conservation genomics

Berthold Heinze
Federal Research Centre for Forests,
Vienna, Austria (BFW);
ProCoGen Work Package Dissemination & Training

This project is financially supported by the European Commission under the 7th Framework Programme

OR: A beginner's guide to 15 000 000 000 (and more...) basepairs in as many trees ...

Berthold Heinze
Federal Research Centre for Forests,
Vienna, Austria (BFW);
ProCoGen Work Package Dissemination & Training

This project is financially supported by the European Commission under the 7th Framework Programme

conifers …

- … harbour some of the longest living organisms
  - *Pinus longaeva*, 5000 years
- … harbour some of the largest single organisms
  - 1500 m³ volume
- … tallest, toughest organisms …

Metuselah bristlecone pine
(www.mnn.com)

General Sherman redwood
(Wikipedia)

ProCoGen

conifers can be very appealing to people …

grove of Rocky Mtn bristlecone pine, *P. aristata*, sacred to Native Americans

Goliath Reserve, Mt. Evans, CO (USA)

ProCoGen

… but conifers are also special in evolution:

- a member of the gymnosperms –
  - term comes from the Greek word for "naked seeds“
  - include conifers, cycads, Ginkgo, and Gnetales
  - stand between ferns and angiosperms in evolution
    - a now extinct member was ancestor to angiosperms
    - are a bit “odd” compared to the angiosperms that we are most familiar with …
    - have evolved before honey bees and other insect pollinators
- conifers are thus „old agers“ –
  - have been around on earth for pretty long time
  - have had lots of time to evolve
  - yet have been slow to evolve … *

* I‘ll come back to that later

„genomics“ - Bu ne?

- the genome of a species has been "sequenced“ = determination of the sequences of one set of chromosomes
- "genome sequence" may be a composite read from the chromosomes of various individuals
- the study of the **global** properties of genomes of related organisms is usually referred to as **genomics**
  - while in genetics, single (or groups of) genes are regarded
- number of base pairs and the number of genes vary widely from one species to another
  - there is only a rough correlation between the two:
  - “C-value paradox”
- the highest known number of genes in any organism:
  - around 60,000 (the protozoan causing trichomoniasis)
  - = almost 3x as many as in the human genome

(Wikipedia)

describing a genome

- measure / determine:
  - chromosome number, karyotype
  - genome size (how much DNA)
  - gene order
  - codon usage bias
    - among multiple codons for same amino acid
  - GC-content
    - percentage of G and C bases
    - opposite of AT-content
    - GC + AT = 100
    - e.g. plant chloroplasts, bacteria – high AT-content

genome sequencing

- determination of nucleotide sequences of whole organisms
  - started in 1970ies: RNA viruses, DNA phages
  - Sanger (dideoxy chain termination) method
  - first complete chloroplast genome – tobacco – 1985
  - first bacterial genome *Haemophilus influenzae* 1995
  - first eukaryotic genome: 16 chromosomes of yeast *Saccharomyces cerevisia* 1995
- other milestones:
  - *Escherichia coli* – 4.6Mb genome size – 1997
  - *Arabidopsis thaliana* 157Mb – first plant – 2000
  - *Homo sapiens* – 3.2Gb – 2001
  - *Populus trichocarpa* 480Mb – first tree – 2006

first conifer genome sequences:

- *Picea abies*
  - 2013 – Swedish consortium/project
- *Picea glauca*
  - 2013 – Canadian consortium/project
- *Pinus taeda*
  - 2013/14 – U.S. consortium/project

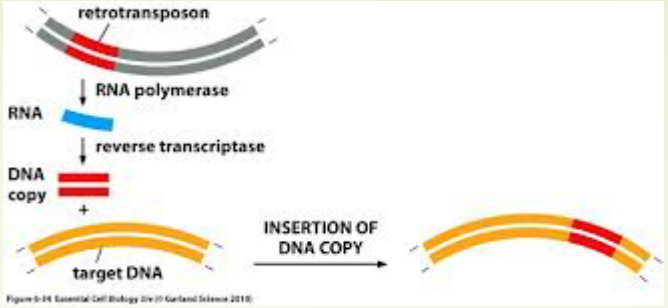
- what does it mean?
  - DNA sequences can be downloaded …
  - there is some description about it
    - identified genes, RNAs, other features …

ProCoGen

genome size varies much more than gene number – why?

- repetitive DNA causes much of the difference
- often, lower and simpler species have less of it
  - *E. coli* – no repetitive DNA
  - *C. elegans, Drosophila* – more non-repetitive DNA (than repetitive)
  - higher eukaryotes often have more repetitive DNA (than non-repetitive)
  - some amphibians and plants have only 20% non-repetitive DNA
- repetitive DNA is comprised of different elements:
  - tandem repeats (satellites, microsatellites)
  - interspersed repeats (transposable elements, some large protein coding gene families, pseudogenes)

transposable elements – retrotransposons - mi ez?



- are transcribed into RNA, then duplicated at other site in the genome
- long terminal repeat (LTRs) and non-LTR elements
- LTRs are similar to viruses (but without envelope)
  - can only act within cells
  - most important source of genome size variation in plants
- Non-LTRs: long / short interspersed elements (LINEs, SINEs); Penelope-like elements
  - LINEs contain transcriptase and endonuclease genes
  - SINEs are smaller than 5000 bp;
    - need interaction with LINEs for their transposition

ProCoGen

genome evolution – what we know fom angiosperms

- duplications play a major role:
  - extension of short tandem repeats
  - duplication of (clusters of) genes
  - duplication of entire chromosomes
  - duplication of entire genomes
- creation new possibilities for evolution:
  - one gene copy retains function,
  - second copy can evolve a new function
  - driver of genetic novelty
- eukaryotic cells have experienced transfer of genetic material from chloroplasts and mitochondria to their nuclear chromosomes
  - major importance in evolution of plants

trad. function

(must) evolve **new function**

ProCoGen

what did we know about conifer genomes before the genomics era?

- often a basic chromosome number of 12
- early observations from allozyme (isoenzyme) studies:
- 1 – high degree of heterozygosity and polymorphism
  - also with each new marker type:
  - RFLP, RAPD and similar, SSR-microsatellites, AFLP, …
- 2 – linkage mapping is possible
  - two isozyme genes on the same chromosome
  - pairs of certain alleles are more frequent in a population than by chance
  - „easy“ because of haploid megagametophytes in seeds:
    - analyse an array of megagametophytes from the same tree
- large overall genome sizes
  - Valkonen et al. 1994 – *P. sylvestris*
    - 1C = 28 pg; 2C = 55 pg
  - Murray 1998 (review):
    - 13 pg (*Metasequoia*) to 63.5 pg (*P. lambertiana)*
    - technical issues in some earlier measurements
    - unclear situation as to intra-specific variation in genome size

further recap: conifer genomes ...

- contain many gene families
  - duplications across various scales.
  - within genes; whole genes; ...
- contain a remarkable number of pseudogenes
- contain genes that are similar to angiosperms
  - but also some unique genes
- contain a remarkable multitude of transposon families
- show evidence of gene sequence conservation
- few genes in large areas of repetitive DNA
  - no mechanism for “easily” getting rid of “junk DNA”?

okay, conifers are similar, but more complex than angiosperms ...

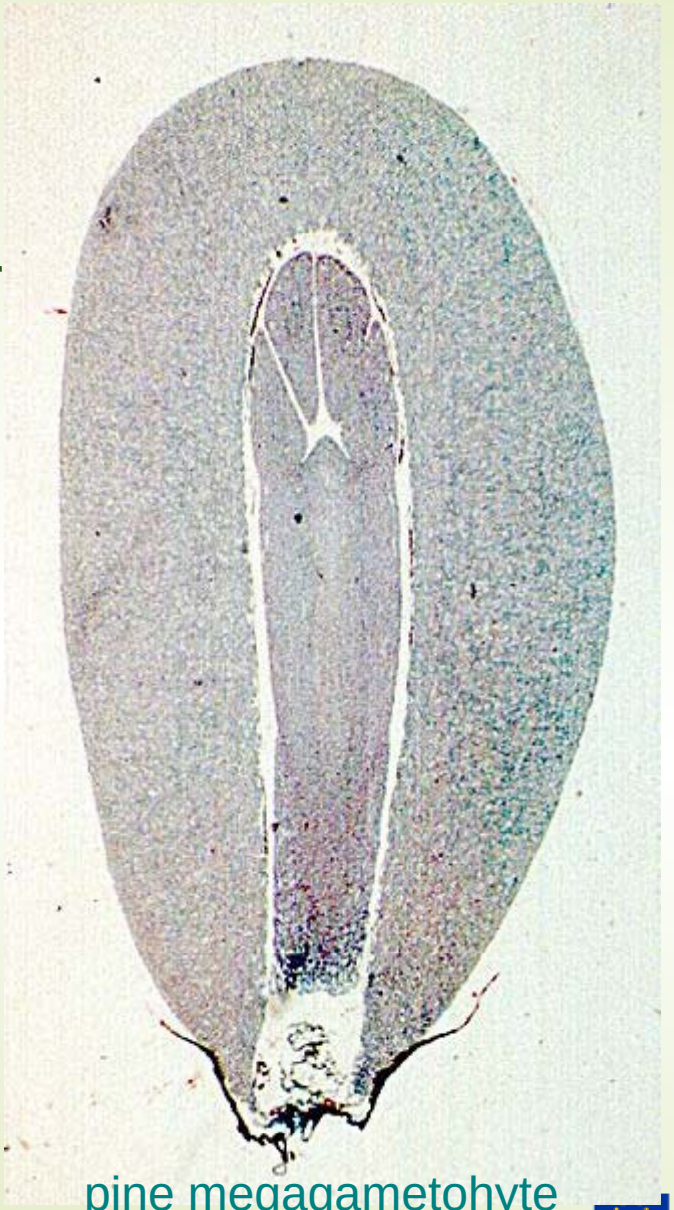
- but how are their typical structures encoded?
  - cones (flowers)
  - tree architecture
  - needle and branch forms
  - wood properties
- how is variation encoded?
- how could they survive for such long evolutionary periods?
- why are there less conifers than angiosperms?
  - which are evolutionary younger as well?
  - do they lack certain „tricks“ of angiosperms?
  - … and still remain evolutionary successful?

that‘s why we study conifer genomics !

ProCoGen

how to study conifer genomics (I): haploid megagametophytes

- seed tissue that surrounds the embryo
- halpoid component of the maother
  - identical to haploid component of the embryo
- can be used for linkage mapping
  - analyse many megagametophytes from a single tree
  - problem – tree must be heterozygous for many of the markers
  - problem – for some species, the size of a single megagametophyte is quite small



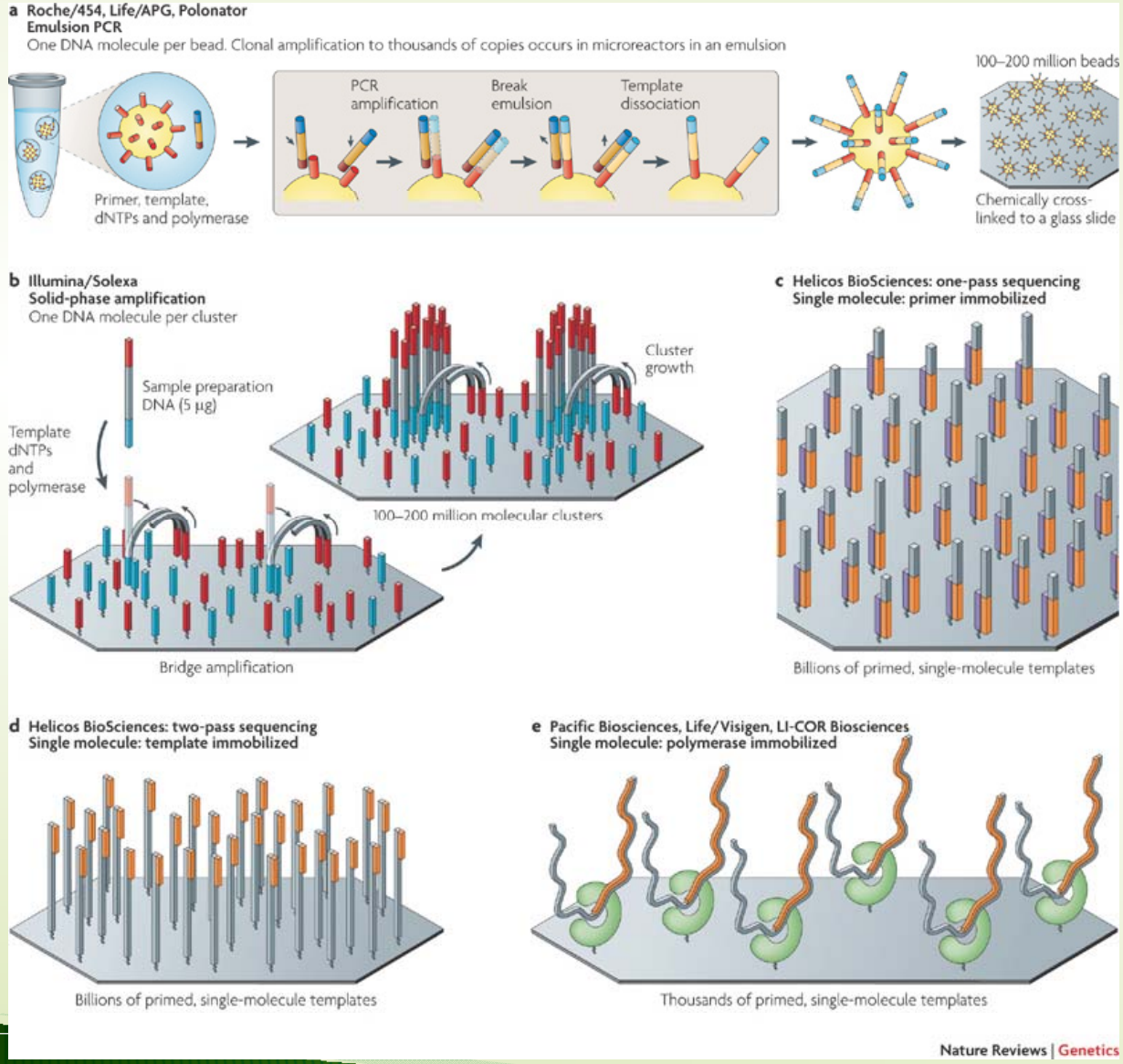
pine megagametohyte (Wikipedia)

ProCoGen

sequence single haploid megagametophytes?

- no heterozygosity problem
- enough DNA for library?

methods (II): „next generation“ techniques



- others include:
- Helicos „Polonator“
- Pacific Biosciences (PacBio)
  - single molecule sequencing
- Ion Torrent
  - electrical charge on a silicon chip
- Oxford Nanopore

ProCoGen

methods (III): bioinformatics

- next gen sequencing and genotyping produces „lots of data“
- traditional (PC) software often not suitable
  - from data download to processing …
  - uncertainties (error percentage) comes into play
- specialist knowledge between biology and computer programming necessary – **bioinformatics**
- often not possible to just download and run software for a given experiment/problem/project
  - needs adjustments – combinations („pipelines“) – own development …

now we know the challenge …

- big, complex genomes
- conifer trees form important ecosystems
- desire to know **how they function „internally“**

- well developed toolbox available on the lab side
- needs equally well developed computer (bioinformatics) toolbox

- now let‘s turn ot Conservation …

ProCoGen

conservation - kas tai?

- first and originally – the conservation of (rare, endangered) species
- additionally – conservation of variants (races, types, …) within a species – genetic conservation
- more generally – conservation of biodiversity
- with all its levels:
  - species
  - genes – genetic diversity within species
  - habitats and ecosystems

- all of this may apply to conifers:

ProCoGen

rare conifer species

- *Pinus torreyana*:
  - ~3000 trees in San Diego, CA (USA); some on an island
  - up from only ~100 trees 100 years ago!

- *Pinus squamata*, Qiaojia pine, 巧家五针松, southern lacebark pine in China
  - only ~20 trees remaining

ProCoGen

rare conifer ecosystems

- “Habitats Directive” of EU
  - Council Directive 92/43/EEC on the Conservation of natural habitats and of wild fauna and flora, 1992
  - **cornerstone of Europe's nature conservation policy**
- lists e.g.:
  - 91. Forests of Temperate Europe:
  - 9120 Atlantic acidophilous beech forests with *Ilex* and sometimes also ***Taxus*** in the shrublayer (*Quercion robori-petraeae* or *Ilici-Fagenion*)
  - 91J0 * *Taxus baccata* woods of the British Isles
  - * Pannonic inland sand dune thicket (*Junipero-Populetum albae)*
  - 91Q0 Western Carpathian calcicolous *Pinus sylvestris forests*
  - 91T0 Central European lichen Scots pine forests

more EC protected conifer forest ecosystems

- **94. Temperate mountainous coniferous forests**
- 9410 Acidophilous *Picea* forests of the montane to alpine levels *(Vaccinio-Piceetea)*
- 9420 Alpine *Larix decidua* and/or *Pinus cembra* forests
- 9430 Subalpine and montane *Pinus uncinata* forests (* if on gypsum or limestone)
- **95. Mediterranean and Macaronesian mountainous coniferous forests**
- 9510 * Southern Apennine *Abies alba* forests
- 9580 * Mediterranean *Taxus baccata* woods
- 9590 * *Cedrus brevifolia* forests *(Cedrosetum brevifoliae)*

rare conifer genetic variants

- „Polish larch“ „*Larix polonica*“:
  - disjunct lowland populations or *L. decidua* in northern Poland and southern Lithuania
- Texas “Lost Pines”:
  - The **Lost Pines Forest** is a 13-mile (21 km) belt of loblolly pines (*Pinus taeda*) in the U.S. state of Texas, unique because it is disjunct from (next stand more than 160 km away, yet closely genetically related to) the vast expanse of pine trees of parts of Texas, Arkansas, Louisiana, and Oklahoma.
  - A portion of the Lost Pines is located inside the boundaries of Bastrop State Park and Buescher State Park, which ensure the trees are protected from development and logging. However, **a large portion of the forest was burned down in the 2011 Bastrop County Complex fire**. (Wikipedia)

conserving genetic variation

- genetic system of conifers is based on high level of genetic variation:
  - many alleles for most genes analysed
  - individual trees are highly heterozygous
  - populations (even small stands) have great variety of genetic variants (alleles)
  - most of the genetic variation is present within individual stands
    - only small part of it is among stands
      - (different alleles, different allele frequencies)
- could be cause or consequence of their lifestyle …
  - long lifetime, overlapping generations, large populations
- … but should be conserved anyway!

so, how can we define „Conifer Conservation Genomics“ …?

- conserving the genetic systems of conifer species in their **entirety** …
- conserving the genomes‘ possibilities to **function** as they should …
- conserving the **adaptation capacity** of conifer species and populations
- conserving the conditions for **further evolution** of conifer genomes …
- **… how can this be done …?**
- find out what is important for :
  - function
  - survival
  - adaptation and evolution

ProCoGen

Conifer Conservation Genomics …

- … a **set of tools** for studying
  - (gene) function
    - gene networks
  - adaptation
  - evolution
- … a **means to improve**
  - conservation of species, genetic variants, habitats
  - understanding of adaptation and evolution

- conserving the capacity of species to survive and evolve further
- for its own purpose
- for breeding, amenity, landscape protection, …

ProCoGen

what we currently do at BFW …

- genetic tests for seed provenance
  - mostly microsatellites/SSRs to establish link between seed orchard or seed stand, and seedlots
- e.g. *Larix decidua* – tests in recent plantations
- marker testing also in
  - *Pinus sylvestris, Abies alba, Taxus baccata, Picea abies, ..*
  - and several hardwood species
- use of „genomic markers“ - SNPs
  - for *Picea abies* populations – ProCoGen WP5
  - for *Larix decidua* seed orchards – Trees4Future project
- plans for „population genomic“ studies in high alpine conifers - *Pinus cembra, Larix decidua*

ProCoGen

what we offer at BFW …

- Trees4Future project – EU funded:
- „Transnational Access“:
- guest visits to our laboratory – several weeks
  - from EU and Associated Countries
- travel and accomodation reimbursed by project
  - BFW receives lab fee per day from the project
- proposal is assessed by committee
- results of the visit are published jointly

- visit
- www.trees4future.eu for details

ProCoGen

AForGeN: Alpine Forest Genomics Network

- informal and open group interested in adaptation to Alpine environments
- meetings each year
- position paper
- promotion of research collaboration

ProCoGen

thank you for you attention!

berthold.heinze@bfw.gv.at
www.procogen.eu

www.trees4future.eu
http://alpforests-gen.fem-environment.eu/

ProCoGen