Artemia Genome

Stephanie De Vos

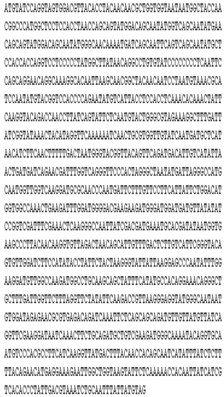
LABORATORY OF AQUACULTURE
ARTEMIA REFERENCE CENTER

Artemia genome workshop
September 2, 2013, Gent

Plant Systems Biology

Overview

1. *Artemia*: genomics model?
2. The *Artemia* genome so far
3. Whole genome sequencing
4. Assembly
5. Annotation
6. First applications
7. What`s next?



Artemia: genomics model?

- Model for toxicity, crustacean metabolism, evolution

- Extremophile: unique genes

  → Osmoregulation (APH-1. *Wang et al., 1999*)

  → Cell cycle (arrest termination in cysts: RSK. *Dai et al., 2008*)

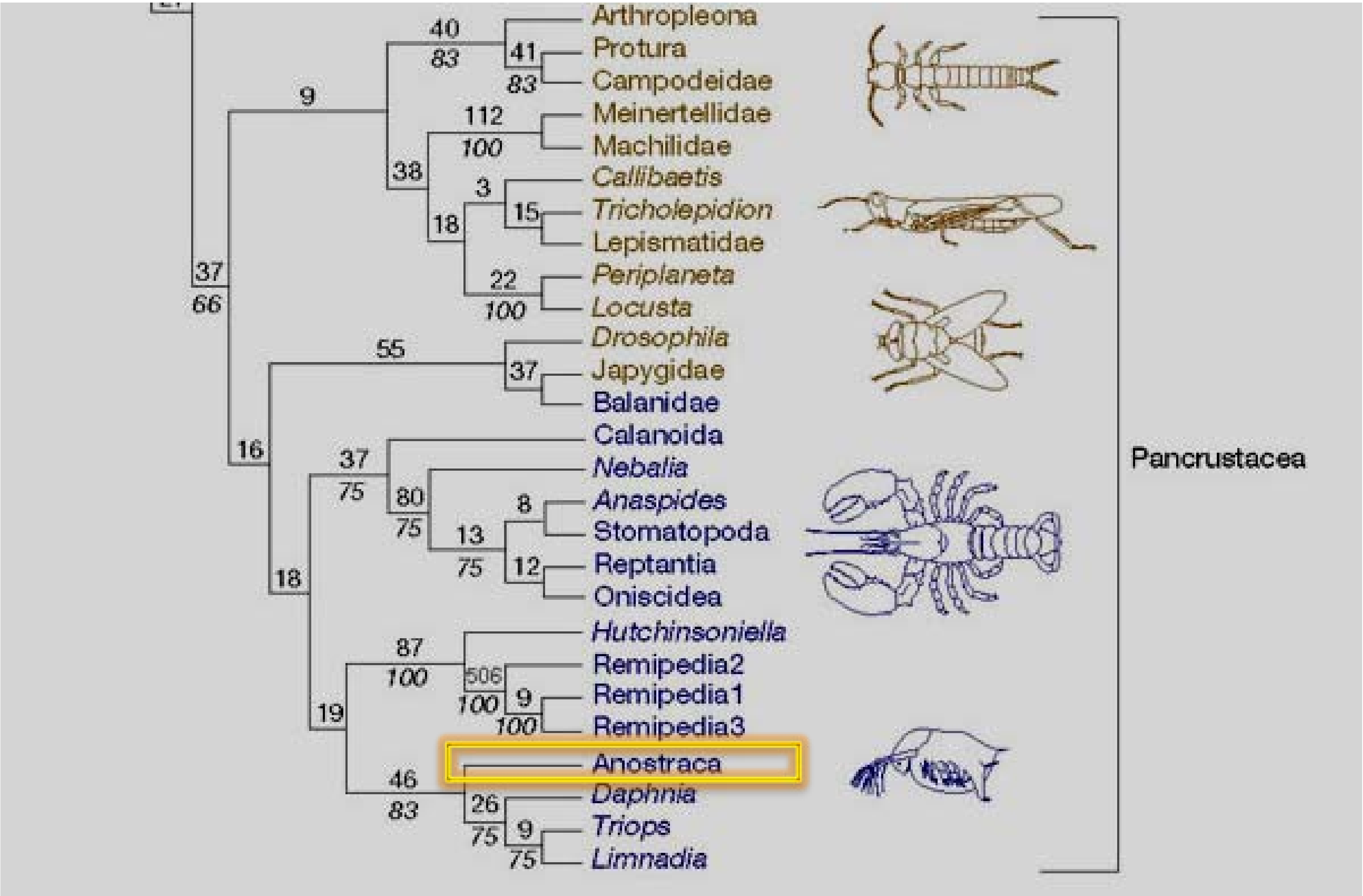
  → Dessication (biostable dried cells: LEA proteins. *Li et al. 2012*)

- Older than *Daphnia*

  → Closer to insect/crustacean branching

  → Better understanding of origin & evolution

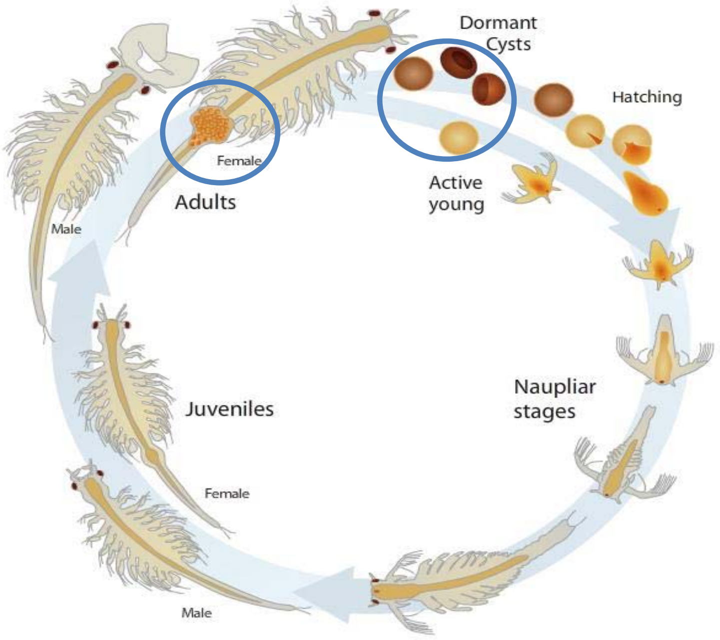
Artemia: genomics model?



Artemia: genomics model?

Pro	Contra
Cysts (Artemia cyst bank)	No cell lines
Large brood	No large collection of inbred strains
Short generation time	Slow uptake of genetic engineering tools
Simple, cheap culture	No mutant collections

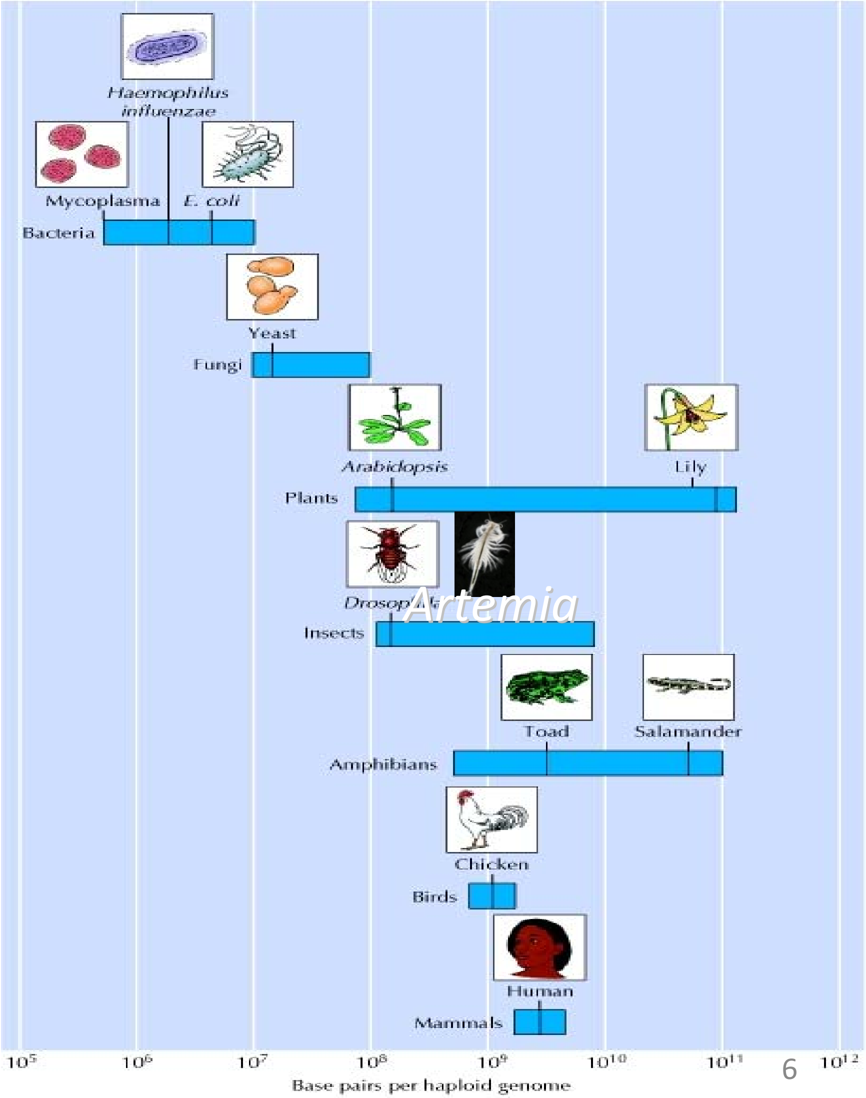
2-4 weeks



The Artemia genome so far

A. franciscana

- mtDNA sequenced
  (Valverde *et al., 1994)*
- 2n = 42
  *(*Abreu-Grobois, *1987)*
  *(*Badaracco et al., *1987)*
- ~ 1 Gb
  *(*De Vos et al., *2013)*
- AFLP-based linkage map
  433 AFLP markers
  *(*De Vos et al., *2013)*



Whole genome sequencing

Materials: *A. franciscana*

Vinh Chau ♀

X

S F Bay ♂

F1

Female pool

65 ♀

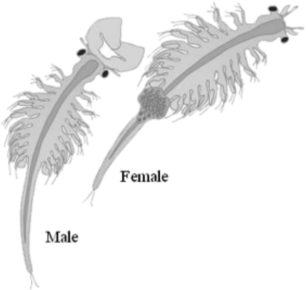
Male pool

55 ♂

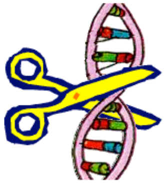
…

Paired-end DNA sequencing
Illumina Hiseq 2000

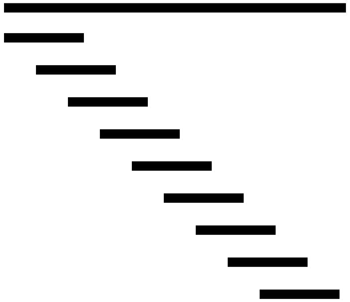
Assembly



Pooled Artemia paired-end



Quality & adaptor trim



Assembly

...

Assembly

…

Scaffold mate-pair

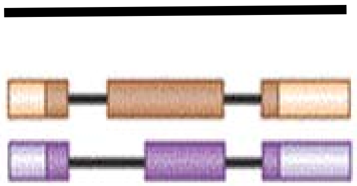
SSpace

Fill gaps

Gapfiller

Remove contamination

NCBI BLAST



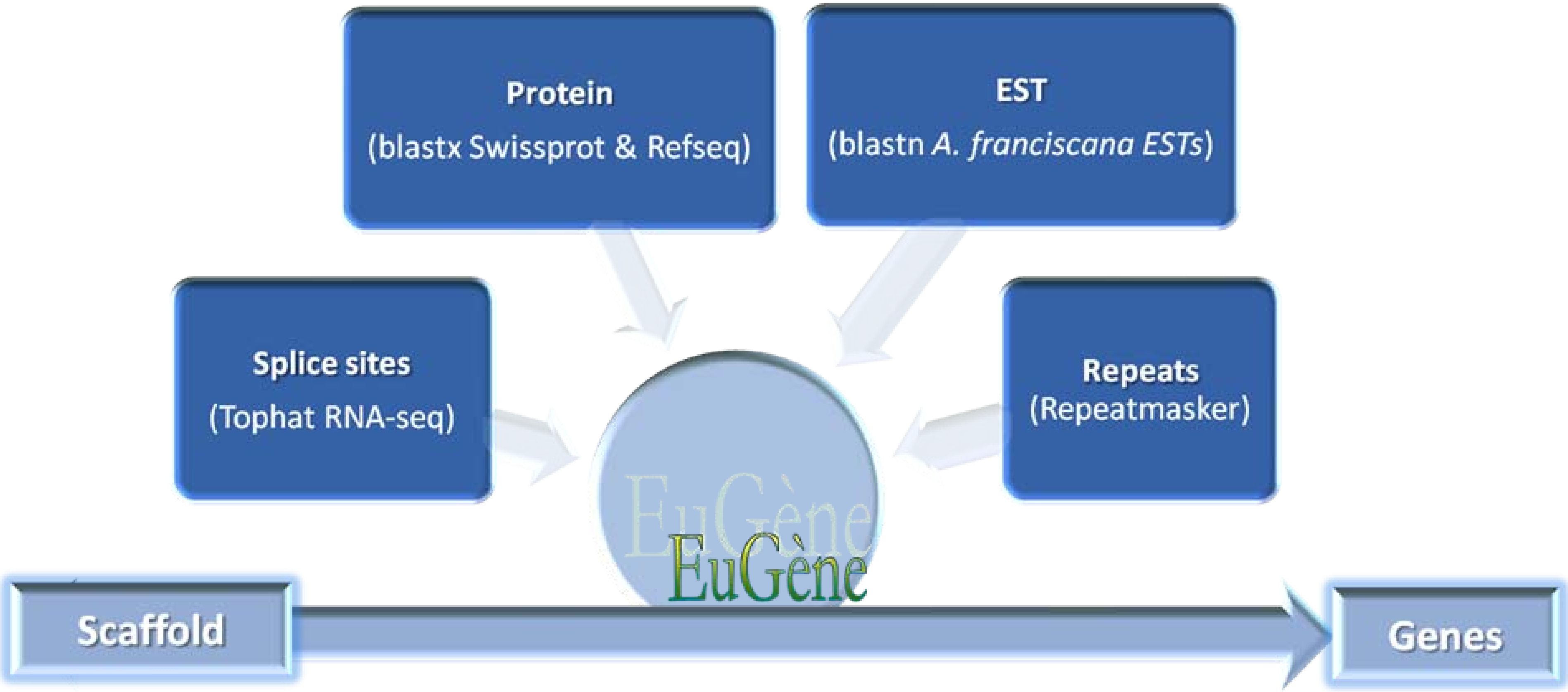
Correct with RNA-seq

CLC bio
Accelerating Scientific Research

Assembly

Coverage	132 X
Scaffolds	176 700
N50	14 800
Assembly size	1.3 Gb
Mean GC content	35 %

Annotation



Protein
(blastx Swissprot & Refseq)

EST
(blastn *A. franciscana ESTs*)

Splice sites
(Tophat RNA-seq)

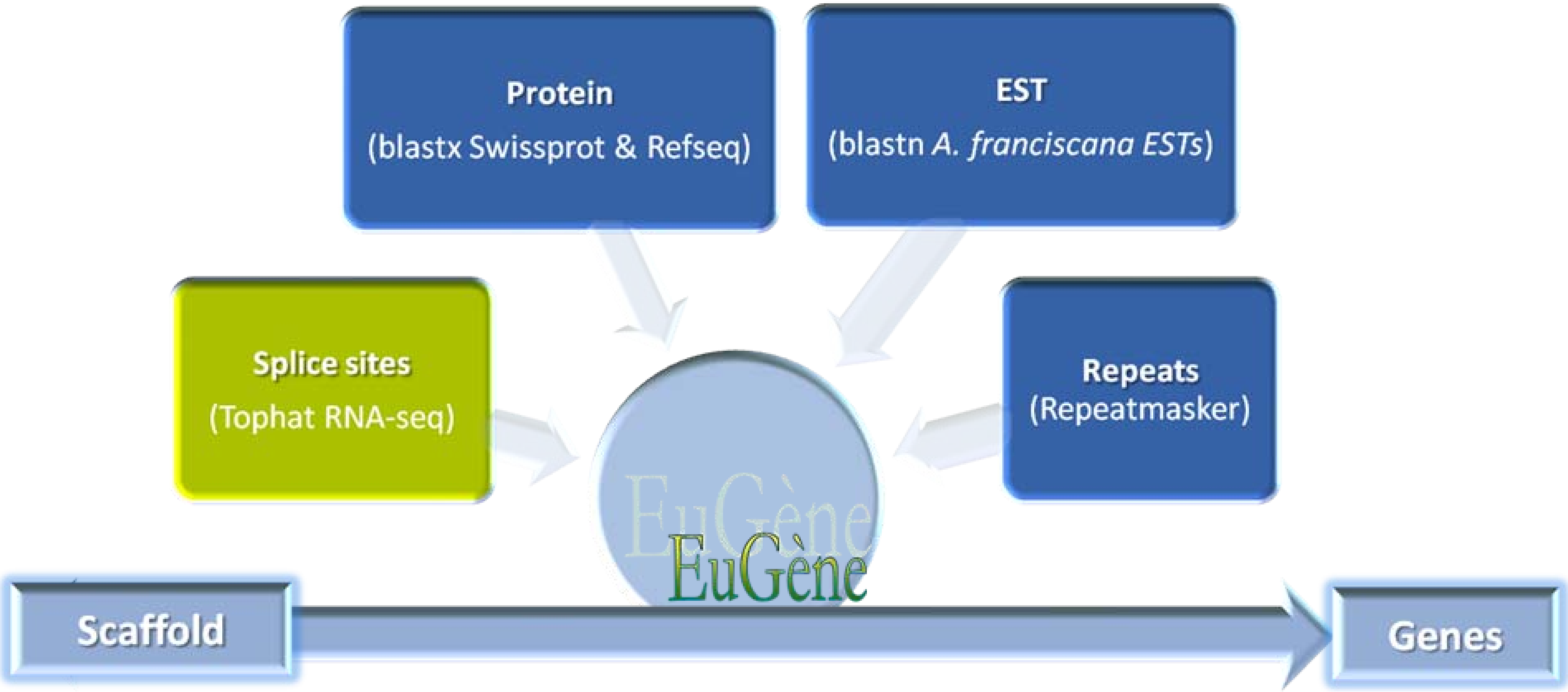
Repeats
(Repeatmasker)

EuGène

Scaffold

Genes

Annotation



Protein
(blastx Swissprot & Refseq)

EST
(blastn *A. franciscana ESTs*)

Splice sites
(Tophat RNA-seq)

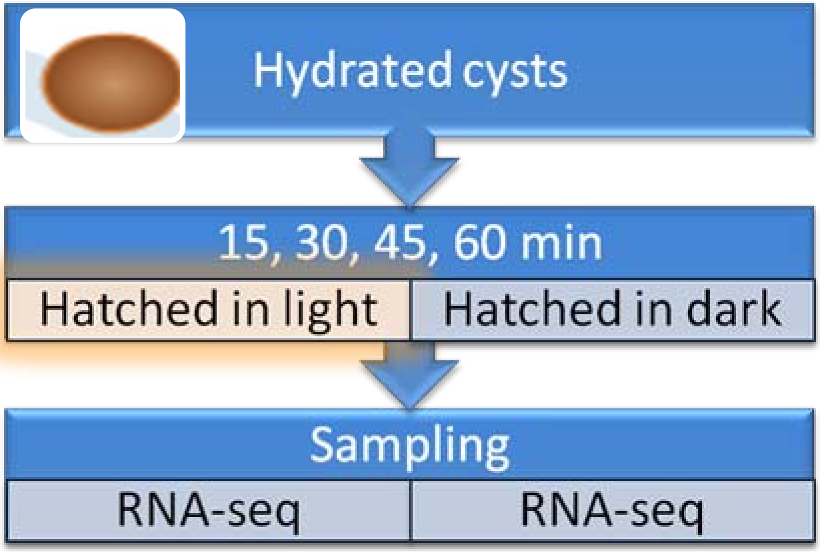
Repeats
(Repeatmasker)

EuGène

Scaffold

Genes

Hatching



Hydrated cysts

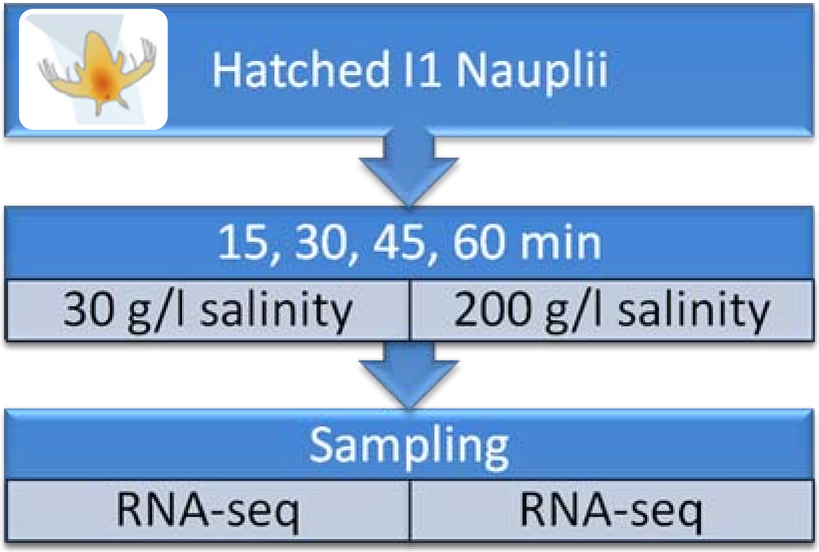
15, 30, 45, 60 min

Hatched in light	Hatched in dark

Sampling

RNA-seq	RNA-seq

Salinity



Hatched I1 Nauplii

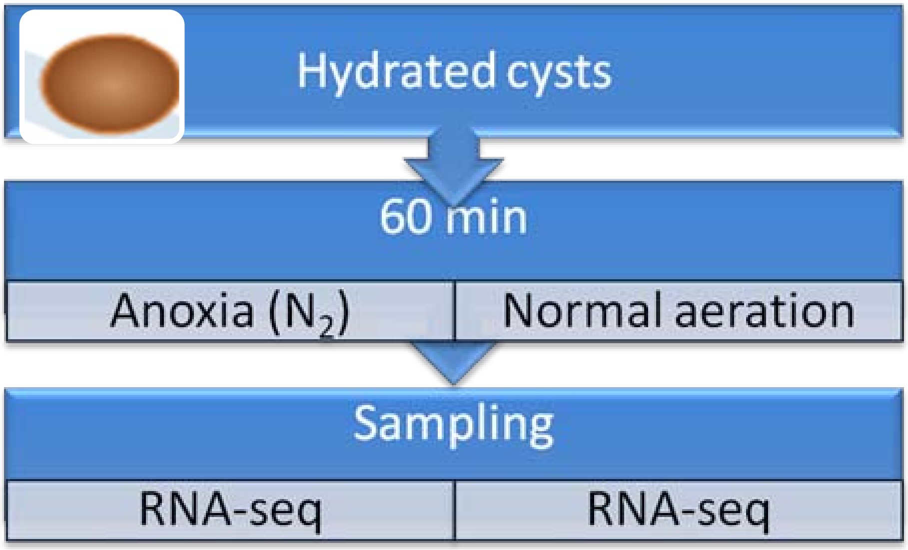
15, 30, 45, 60 min

30 g/l salinity	200 g/l salinity

Sampling

RNA-seq	RNA-seq

Anoxia



Hydrated cysts

60 min

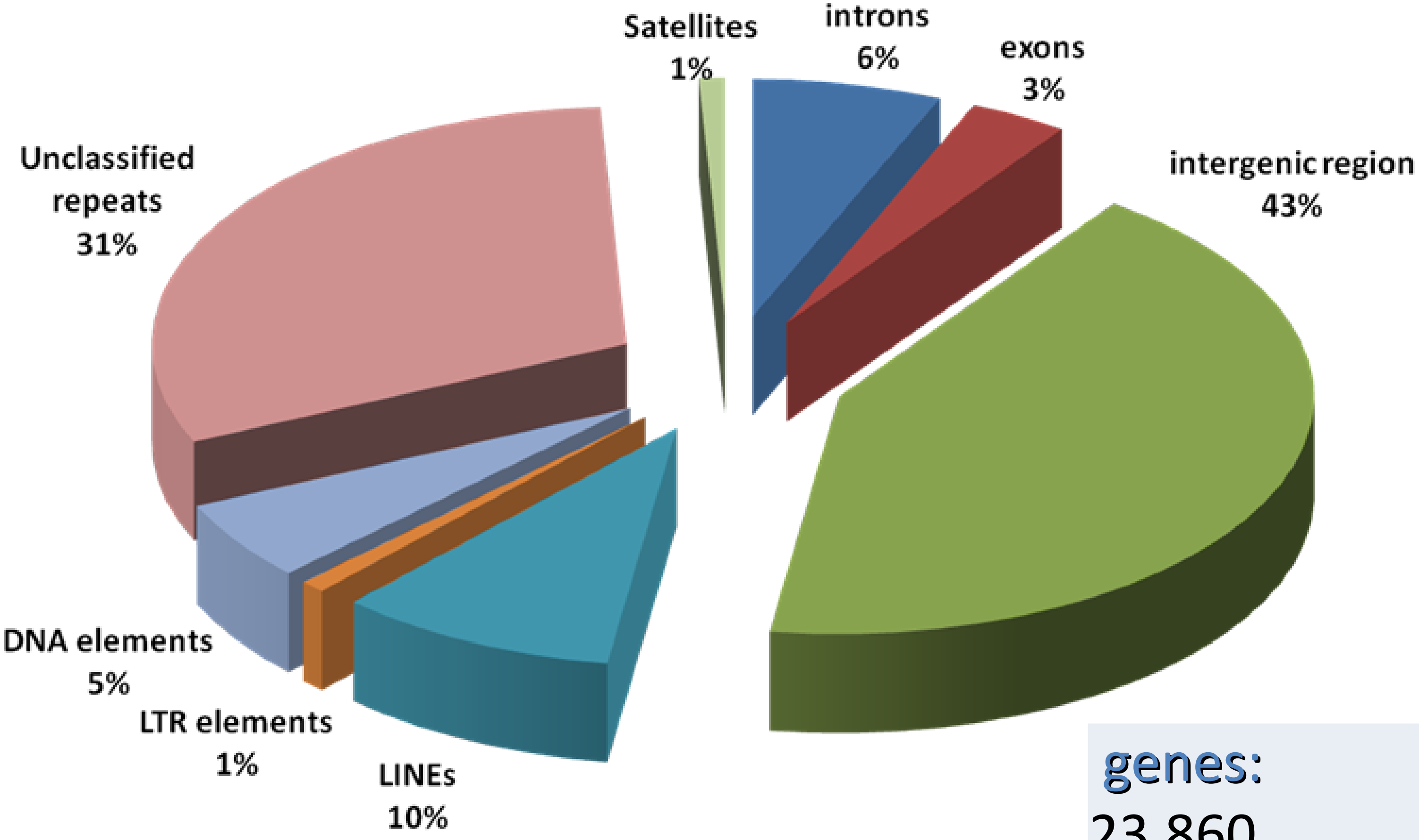
Anoxia (N_2)	Normal aeration

Sampling

RNA-seq	RNA-seq

Annotation

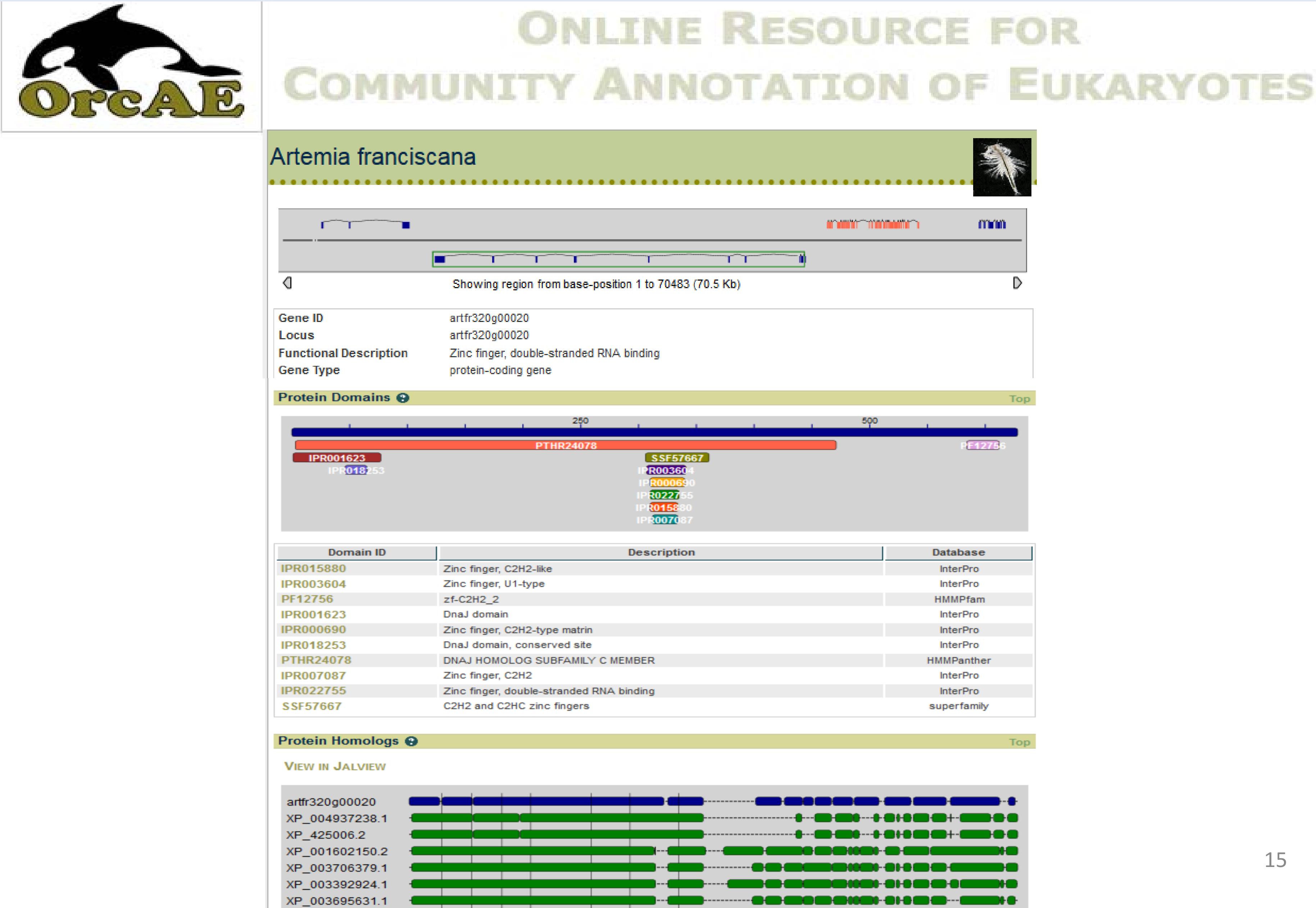
Components of the *Artemia* genome



genes:
23,860

av. length.genes:
1,171 bp

Annotation



OrcAE

ONLINE RESOURCE FOR COMMUNITY ANNOTATION OF EUKARYOTES

Artemia franciscana

Showing region from base-position 1 to 70483 (70.5 Kb)

Gene ID	artfr320g00020
Locus	artfr320g00020
Functional Description	Zinc finger, double-stranded RNA binding
Gene Type	protein-coding gene

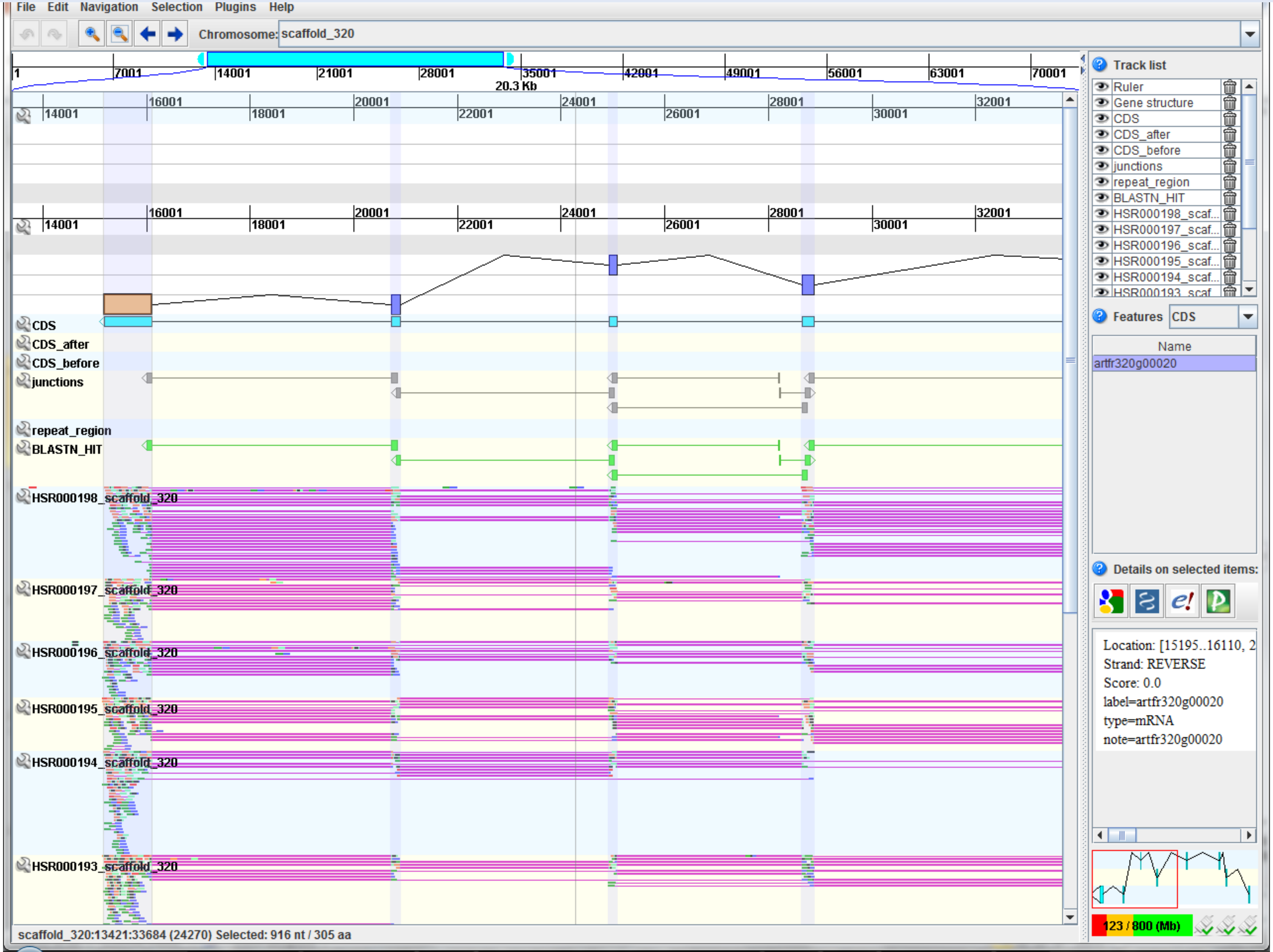
Protein Domains Top

Domain ID	Description	Database
IPR015880	Zinc finger, C2H2-like	InterPro
IPR003604	Zinc finger, U1-type	InterPro
PF12756	zf-C2H2_2	HMMPfam
IPR001623	DnaJ domain	InterPro
IPR000690	Zinc finger, C2H2-type matrin	InterPro
IPR018253	DnaJ domain, conserved site	InterPro
PTHR24078	DNAJ HOMOLOG SUBFAMILY C MEMBER	HMMPanther
IPR007087	Zinc finger, C2H2	InterPro
IPR022755	Zinc finger, double-stranded RNA binding	InterPro
SSF57667	C2H2 and C2HC zinc fingers	superfamily

Protein Homologs Top

VIEW IN JALVIEW

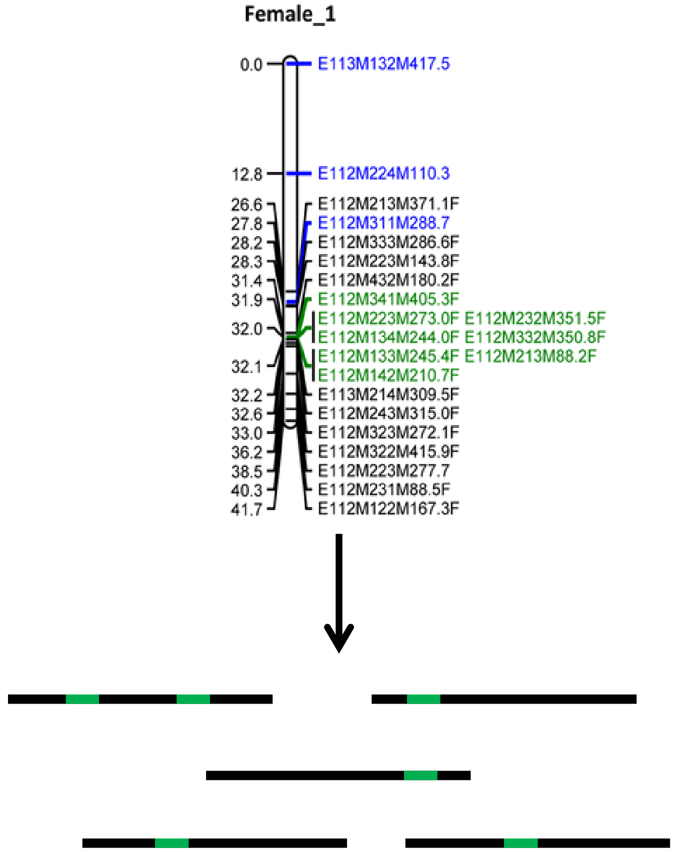
Annotation



Location: [15195..16110, 2
Strand: REVERSE
Score: 0.0
label=artfr320g00020
type=mRNA
note=artfr320g00020

scaffold_320:13421:33684 (24270) Selected: 916 nt / 305 aa

Characterizing the sex-determining region

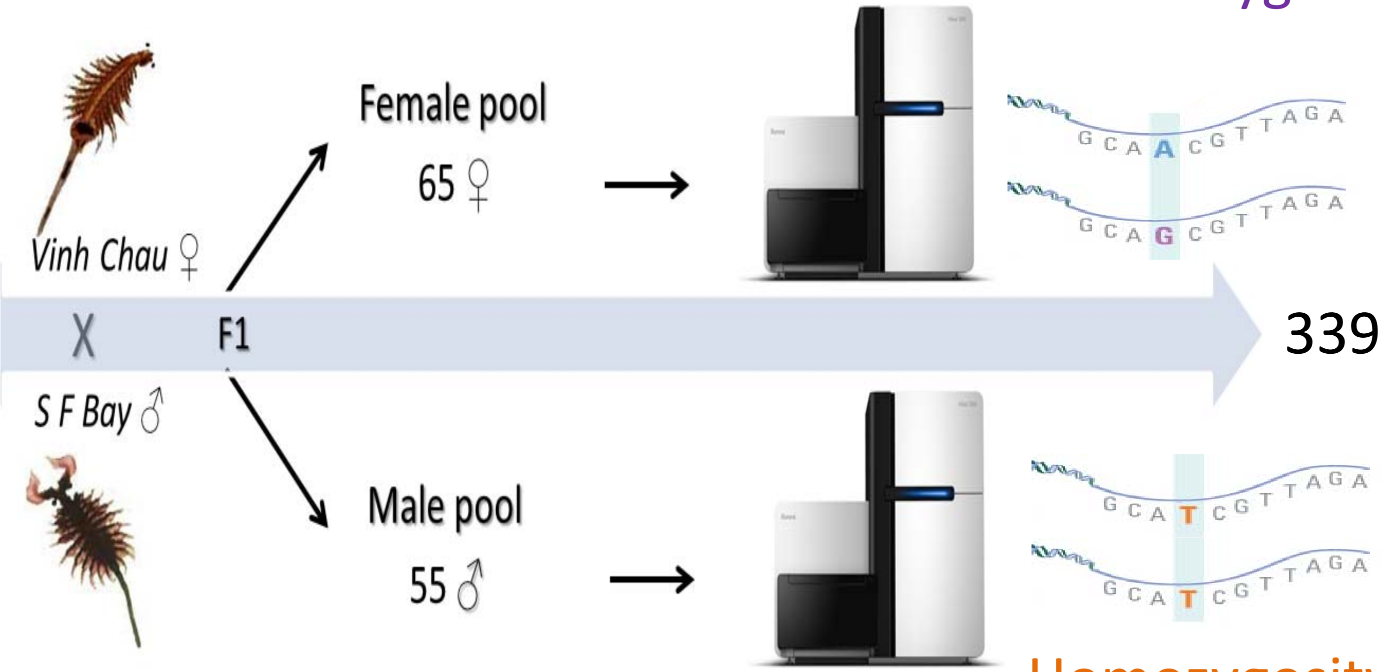


Scaffolds carrying candidate sex-determining genes

Characterizing the sex-determining region

BSA by GWS

Heterozygosity (ZW)

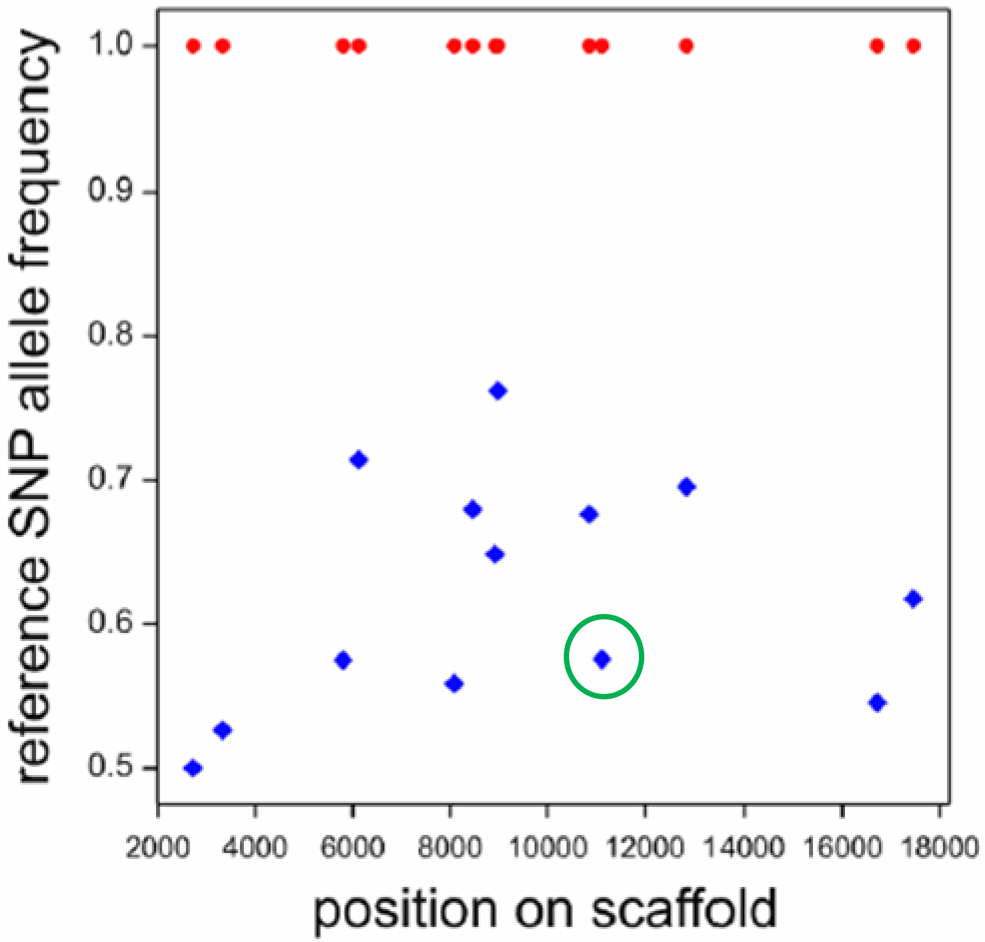


339 scaffolds identified

Homozygosity (ZZ)

Principle: SNP allele frequencies should differ significantly in the sex-determining region

Characterizing the sex-determining region



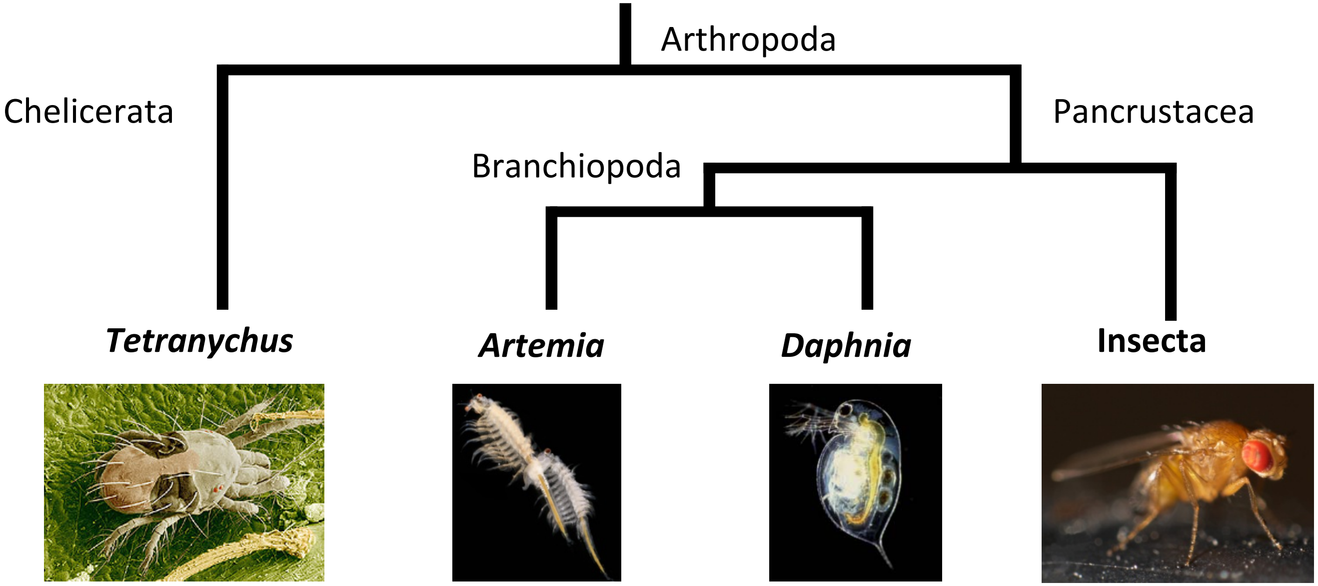
SNPs from Male reads

SNPs from Female reads

On sex-linked marker

Annotation of the nuclear receptors (NR)

- Transcription factors
- Embryonic development, cell differentiation, reproduction, moulting and metamorphosis, …



Arthropoda

Chelicerata

Pancrustacea

Branchiopoda

Tetranychus

Artemia

Daphnia

Insecta

30 NRs

?

25 NRs

19-22 NRs

Annotation of the nuclear receptors (NR)

Preliminary NR annotation results

- Possible *Artemia*-specific NR subfamily expansions (lineage-specific duplications)
- Some *Daphnia-specific* NRs missing in *Artemia*
- Two Ecdysone receptors (cfr *Daphnia*)
- Preliminary results: Pseudogenes? Redundancy?

→ **A distinct set of NRs in *Artemia*, very different from *Daphnia pulex***

What`s next?

NGS trade-off

Pro

- Cheap
- Many reads
- Fast

Contra

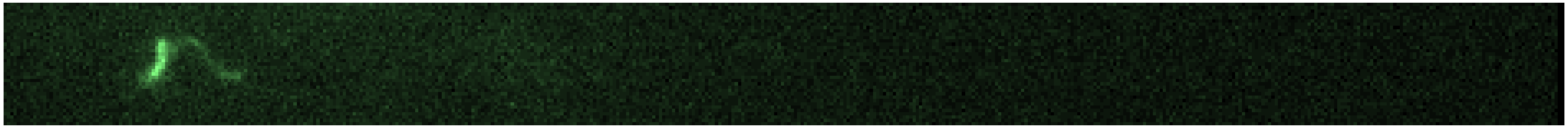
- More errors
- Shorter: harder assembly
  more fragmented assembly
- Harder annotation

What`s next?

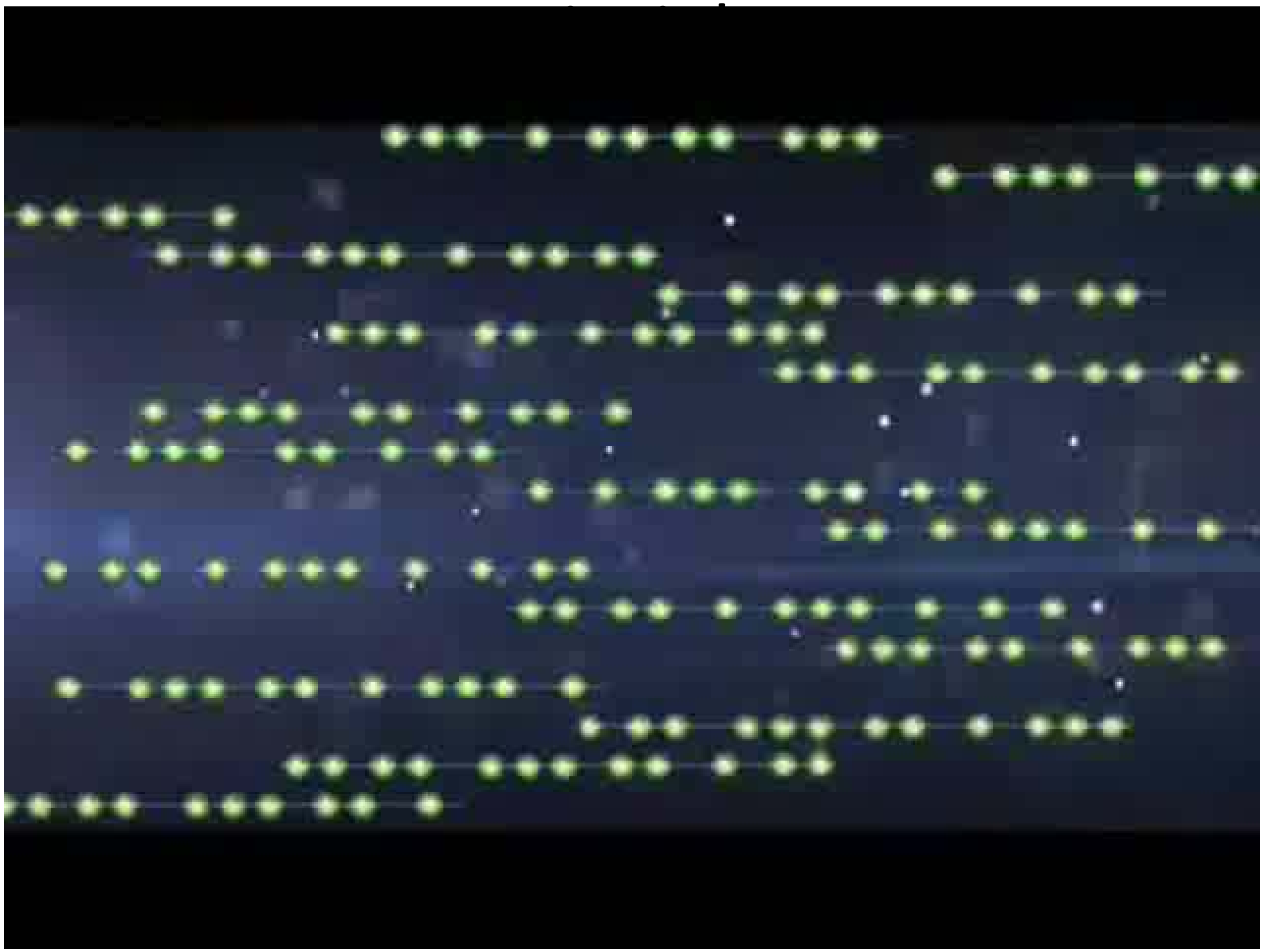
BIONANO
GENOMICS

Visualization of extremely long DNA molecules in nanochannels
"Single-molecule" imaging

→ Finished *de novo* genome of higher quality



What`s next?



Acknowledgements

UNIVERSITEIT GENT

prof. dr. ir. Peter Bossier

dr. Gilbert Van Stappen

prof. dr. Patrick Sorgeloos

prof. dr. ir. Guy Smagghe

dr. ir. Olivier Christiaens

Plant Systems Biology

prof. dr. ir. Marnik Vuylsteke

bioinformatics Gent

Stephane Rombauts