

1,000 Insect Transcriptomes

Introducing the 1KITE project, taking next steps in phylogenomics and what else can be done with transcriptomic data

Karen Meusemann
on behalf of the 1KITE Consortium



CSIRO Ecosystem Sciences, Australian National Insect Collection, Canberra, Australia

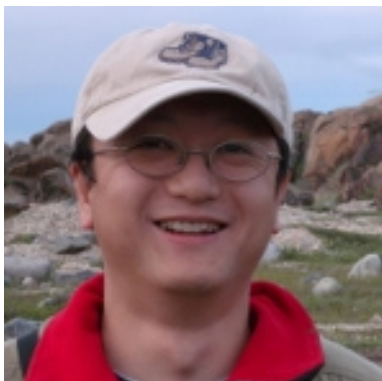
Phylomania, Hobart, Tasmania, November 08, 2013



Motivation

"We want to understand the evolution of insects in all of its aspects, using the results of the most advanced approaches in genomic and morphological research."

Speakers



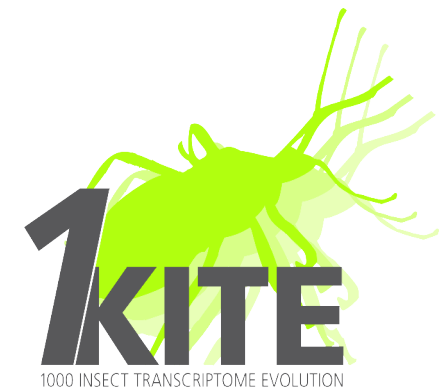
Xin Zhou,
BGI Shenzhen



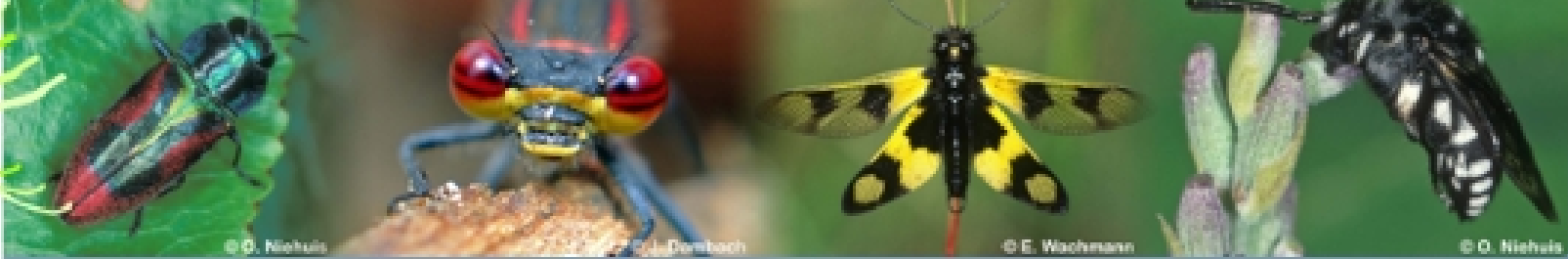
Karl Kjer,
Rutgers, NJ



Bernhard Misof,
ZMB, ZFMK Bonn



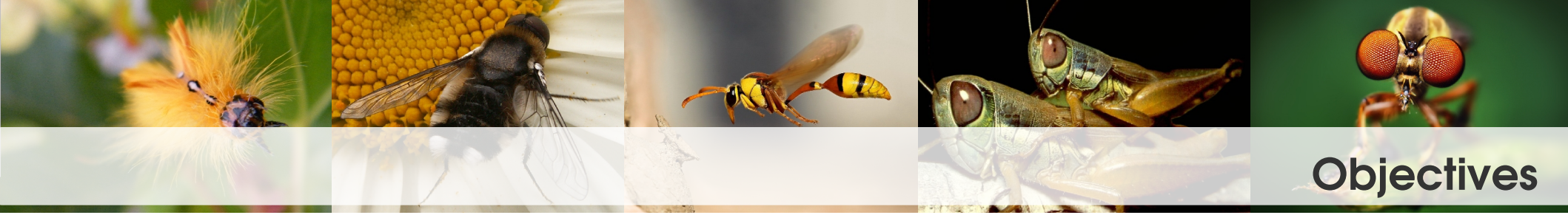
& 1KITE team

[Home](#)[Collaborators](#)[Responsibilities](#)[Subprojects](#)[News](#)[Wiki \(intern\)](#)

Collaborators

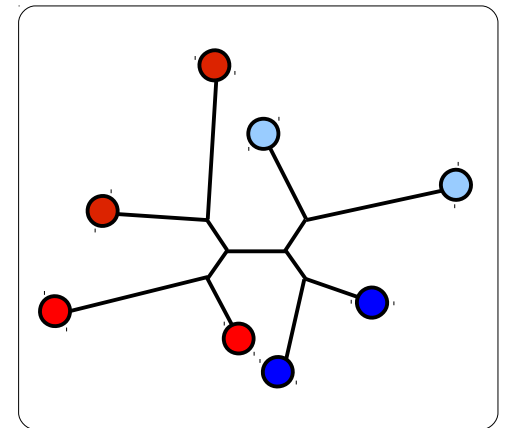


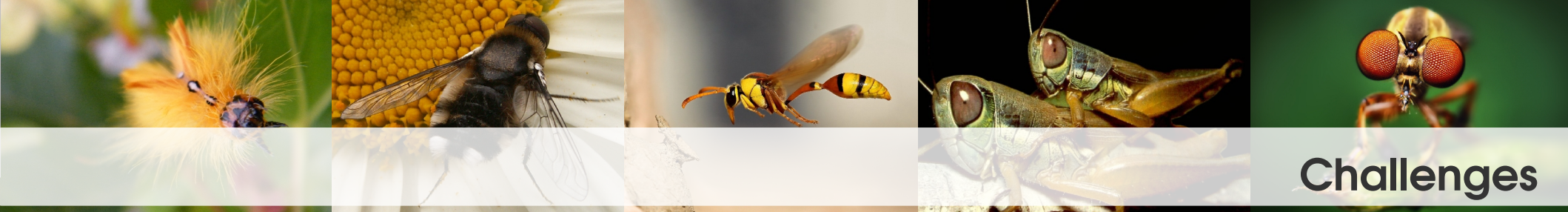
The 1KITE Consortium and Collaborators involve more than 120 scientists (16 nations) from Molecular Biology, Morphology, Systematics, Paleontology, Embryology, Bioinformatics, and Scientific Computing and beyond.



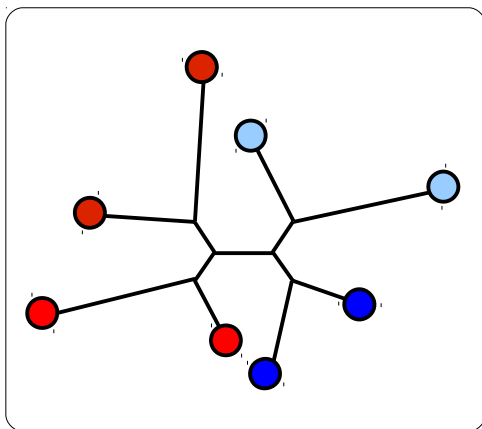
Objectives

- 1 - Phylogenetic relationships
- 2 - Understanding phenotypic evolution
- 3 - Can we push molecular analyses to their limits and what do we gain?

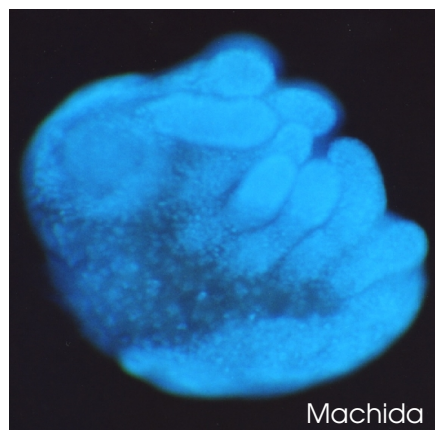
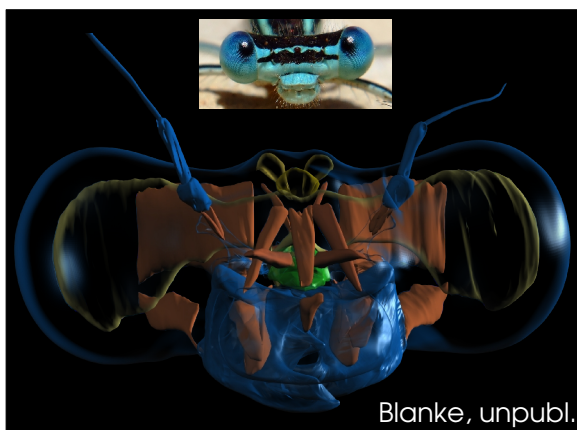




Challenges



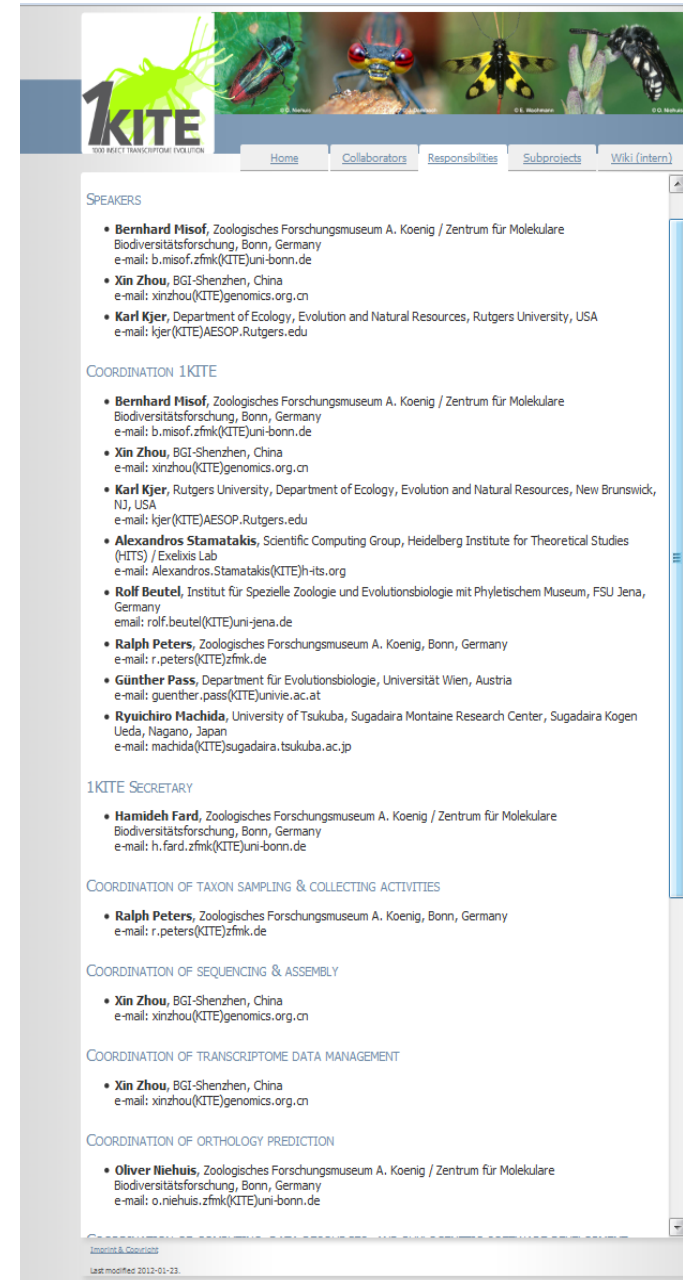
- Infer robust phylogeny
- Understand evolution
- Explain biodiversity



Coordination of

- taxon sampling and collecting activities
- data management
- sequencing and assembly
- orthology prediction
- computing, data resources, and phylogenetic software development
- phylogenetic workflow
- plausibility assessment
- taxonomic subprojects

see www.1kite.org



1KITE
100 REAL TRANSFORME EVOLUTION

Home Collaborators Responsibilities Subprojects Wiki (intern)

SPEAKERS

- **Bernhard Misof**, Zoologisches Forschungsmuseum A. Koenig / Zentrum für Molekulare Biodiversitätsforschung, Bonn, Germany
e-mail: b.misof.zfmk@1kite.uni-bonn.de
- **Xin Zhou**, BGI-Shenzhen, China
e-mail: xinzhou@1kite.genomics.org.cn
- **Karl Kjer**, Department of Ecology, Evolution and Natural Resources, Rutgers University, USA
e-mail: kjer@1kite.AESOP.Rutgers.edu

COORDINATION 1KITE

- **Bernhard Misof**, Zoologisches Forschungsmuseum A. Koenig / Zentrum für Molekulare Biodiversitätsforschung, Bonn, Germany
e-mail: b.misof.zfmk@1kite.uni-bonn.de
- **Xin Zhou**, BGI-Shenzhen, China
e-mail: xinzhou@1kite.genomics.org.cn
- **Karl Kjer**, Rutgers University, Department of Ecology, Evolution and Natural Resources, New Brunswick, NJ, USA
e-mail: kjer@1kite.AESOP.Rutgers.edu
- **Alexandros Stamatakis**, Scientific Computing Group, Heidelberg Institute for Theoretical Studies (HITS) / Exelixis Lab
e-mail: Alexandros.Stamatakis@1kite-hits.org
- **Rolf Beutel**, Institut für Spezielle Zoologie und Evolutionsbiologie mit Phyletischem Museum, FSU Jena, Germany
email: rolf.beutel@1kite.uni-jena.de
- **Ralph Peters**, Zoologisches Forschungsmuseum A. Koenig, Bonn, Germany
e-mail: r.peters@1kite.zfmk.de
- **Günther Pass**, Department für Evolutionsbiologie, Universität Wien, Austria
e-mail: guenther.pass@1kite.univie.ac.at
- **Ryuichiro Machida**, University of Tsukuba, Sugadaira Montaine Research Center, Sugadaira Kogen Ueda, Nagano, Japan
e-mail: machida@1kite.sugadaira.tsukuba.ac.jp

1KITE SECRETARY

- **Hamideh Fard**, Zoologisches Forschungsmuseum A. Koenig / Zentrum für Molekulare Biodiversitätsforschung, Bonn, Germany
e-mail: h.fard.zfmk@1kite.uni-bonn.de

COORDINATION OF TAXON SAMPLING & COLLECTING ACTIVITIES

- **Ralph Peters**, Zoologisches Forschungsmuseum A. Koenig, Bonn, Germany
e-mail: r.peters@1kite.zfmk.de

COORDINATION OF SEQUENCING & ASSEMBLY

- **Xin Zhou**, BGI-Shenzhen, China
e-mail: xinzhou@1kite.genomics.org.cn

COORDINATION OF TRANSCRIPTOME DATA MANAGEMENT

- **Xin Zhou**, BGI-Shenzhen, China
e-mail: xinzhou@1kite.genomics.org.cn

COORDINATION OF ORTHOLOGY PREDICTION

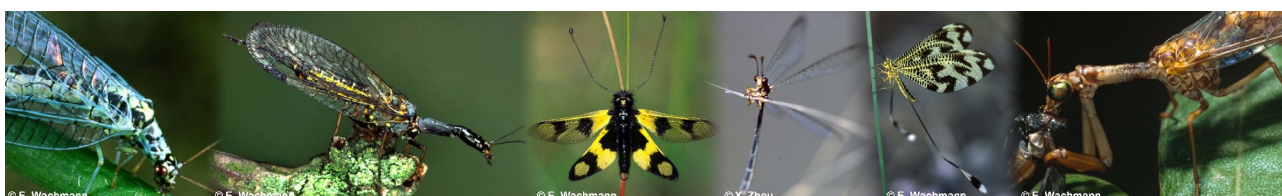
- **Oliver Niehuis**, Zoologisches Forschungsmuseum A. Koenig / Zentrum für Molekulare Biodiversitätsforschung, Bonn, Germany
e-mail: o.niehuis.zfmk@1kite.uni-bonn.de

Terms & Conditions
last modified 2012-01-23

Taxonomic Subprojects

11 Subprojects

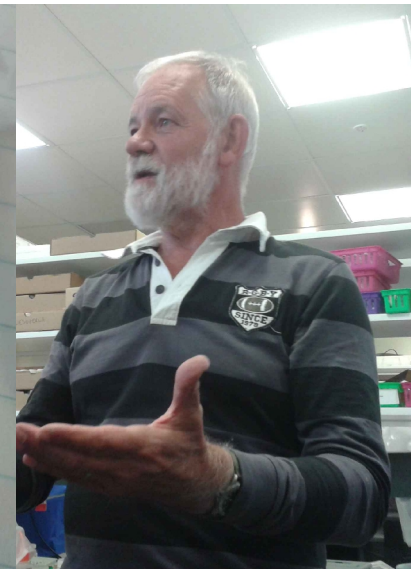
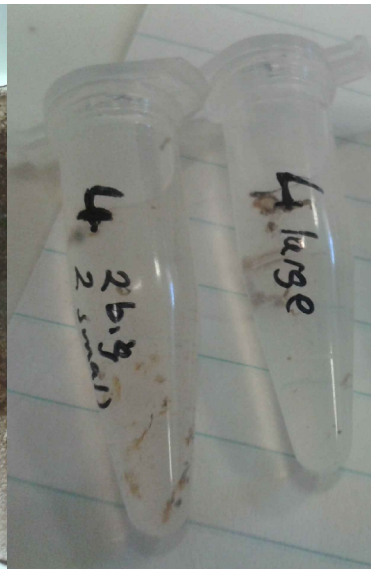
- Basal Hexapods (109)
- TransOdonata (120)
- Polyneoptera (131)
- Dictyoptera (96)
- Paraneoptera (72)
- Hymenoptera (250)
- Neuropterida (95)
- Coleoptera (62)
- Trichoptera (60)
- Lepidoptera (70)
- Antliophora (124)



Collecting more than 1,000 Insect Species



garden springtail *Bourletiella hortensis*, 06/2011



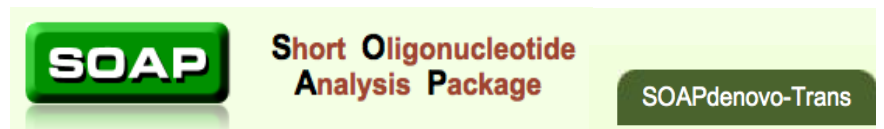
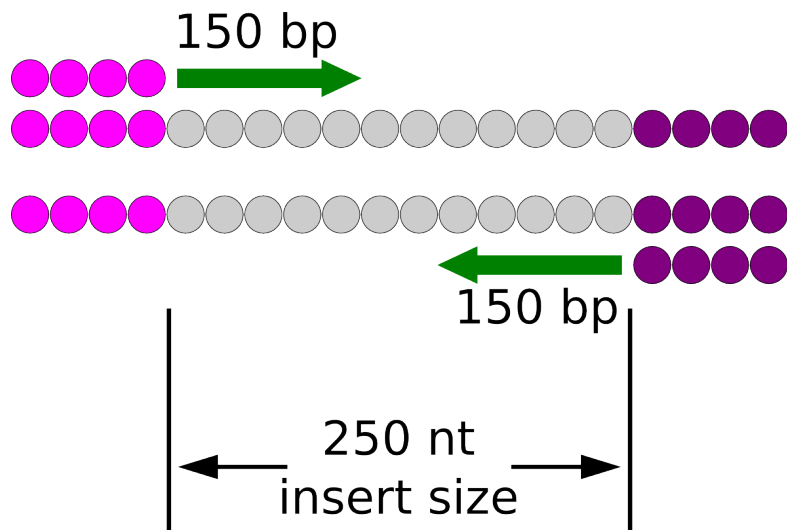
enigmatic mecopterans: *Nannochorista*, 10/2013



Sequencing and Assembly



- 250 bp insert-size libraries
- 150 bp paired-end sequencing
- 2.5 GB raw data per species
- ILLUMINA HiSeq 2000 sequencing
- > 5,000 full length genes

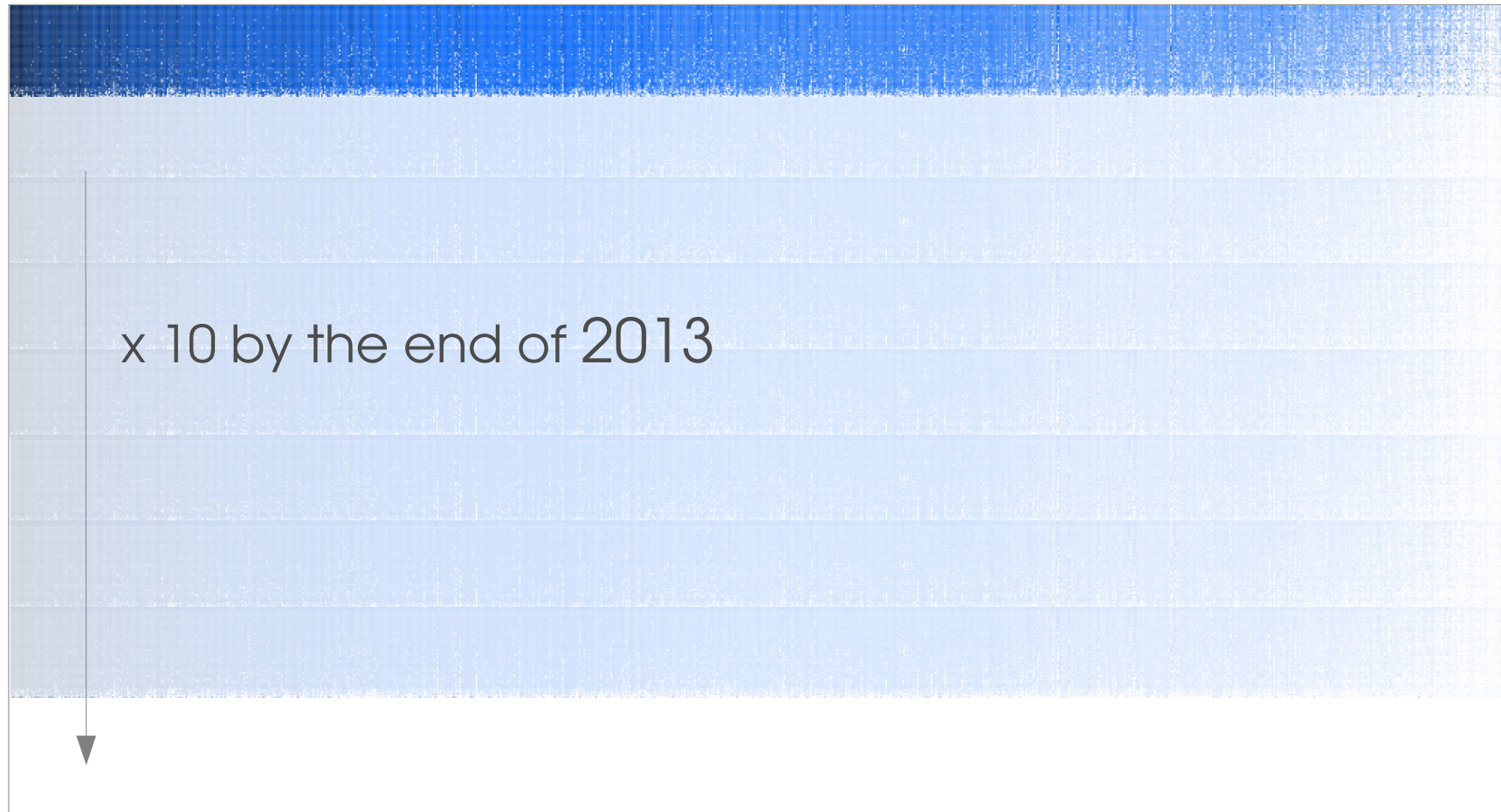


av. contig #	= 92533
av. scaffold #	= 14569
av. N50	= 629bp

Struck et al. (2011)



TKITE,
(in prep.)



Phylogenetic Workflow

- orthology prediction and assignement of transcripts
- annotation, alignment, refinement and alignment masking
- identification of protein domains
- selection of a „decisive“ dataset (reduce missing data)
- sophisticated partitioning (coll. Rob Lanfear, Brett Calcott, ANU)
- exploration of datasets (coll. Lars Jermiin & group)
- evaluate and reduce heterogeneity (coll. Lars Jermiin & group)
- tree reconstruction, dating
- **Four-cluster Likelihood mapping** (test single nodes, detect „hidden“ signal)

Specific Hypothesis Testing - FcLM

- single splits – is there alternative signal present?
- FcLM implemented in RAxML
- new mapping tool
- are the results affected by non-phylogenetic signal

Four-cluster Likelihood Mapping - Specific Hypothesis Testing

Meusemann *et al.* 2010

A Phylogenomic Approach to Resolve the Arthropod Tree of Life

v. Reumont *et al.* 2012

Kar
Pat
Arn
Mis
Zoo
Scif
Bio
Cen
of Ve
Dep
Dep
Reg
Dep
The
Cor
Asso

MOLECULAR BIOLOGY AND EVOLUTION

MBE Advance Access published December 20, 2011

Example: Basal Hexapods: Protura + Diplura (Nonoculata) versus Collembola + Protura (Ellipura)

Bjoern M. von Reumont,^{1*} Ronald A. Jenner,² Matthew A. Wills,³ Emiliano Dell’Ampio,⁴ Günther Pass,⁴ Ingo Ebersberger,⁵ Benjamin Meyer,⁶ Stefan Koenemann,⁷ Thomas M. Iliffe,⁸ Alexandros S. Antoniou,⁹ and Heinrich Schulze¹

¹Zoologisches Forschungsmuseum Alexander Koenig, Zentrum für Molekulare Biodiversitätsforschung, Bonn, Germany

²Department of Zoology, The Natural History Museum, London, United Kingdom

³Department of Biology and Biochemistry, University of Exeter, Exeter, United Kingdom

⁴Department of Evolutionary Biology, University of Vienna, Vienna, Austria

⁵Center for Integrative Animal Research, Institute for Molecular Evolution and Systematics, University of Veterinary Medicine Vienna, Vienna, Austria

⁶Biozentrum Grindel und Zoologisches Museum, Universität Hamburg, Hamburg, Germany

⁷Section Biology, Science and Technology, University of Siegen, Siegen, Germany

⁸Department of Marine Biology, Texas A&M University at Galveston

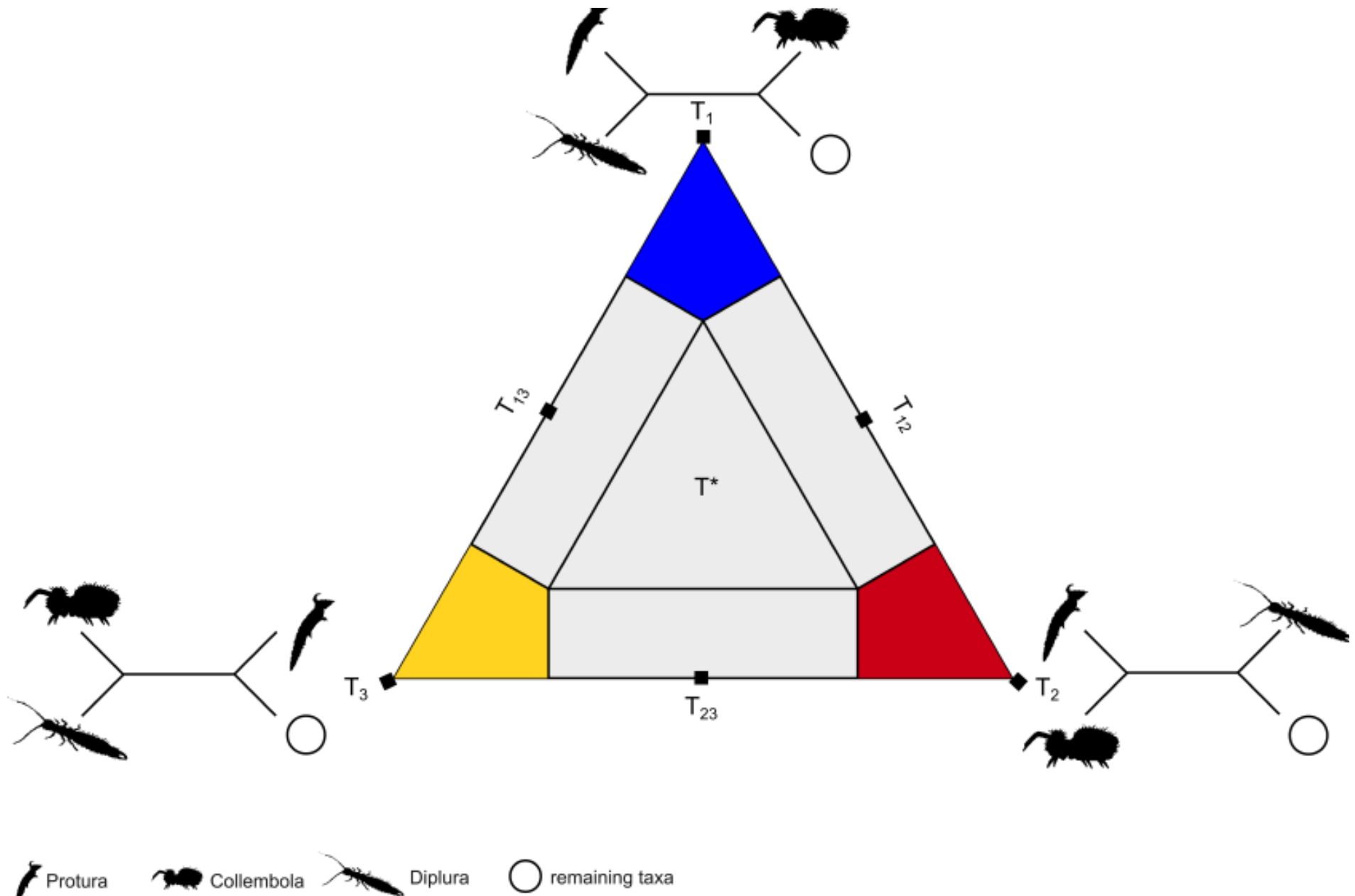
⁹The Exelixis Lab, Scientific Computing Group, Heidelberg Institute for Theoretical Studies, Heidelberg, Germany

*Corresponding author: E-mail: bmvr@arcor.de.

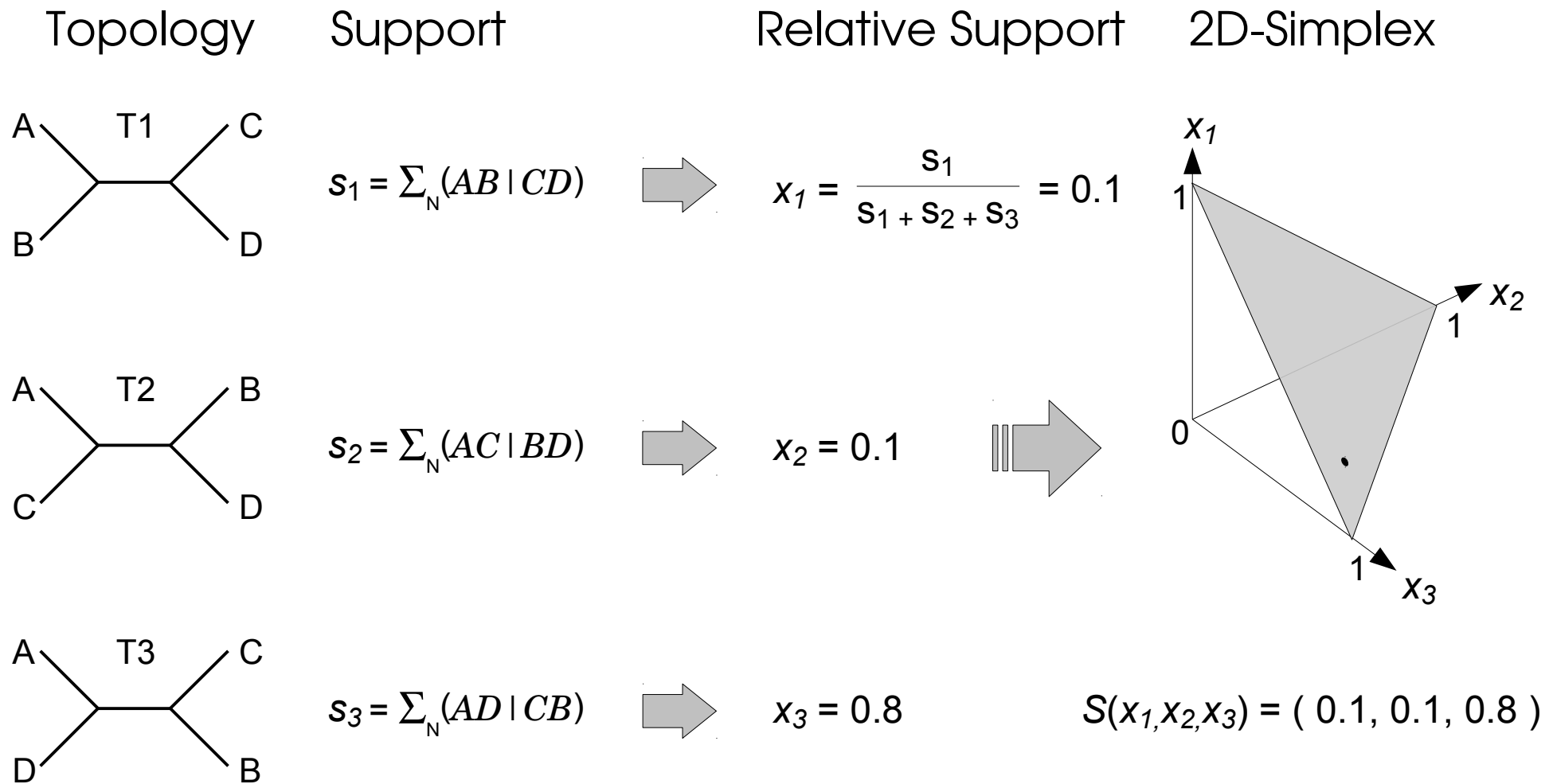
Associate editor: Billie Swalla

Downloaded from <http://ml.oup.com/>
Research article

Four-cluster Likelihood Mapping - Specific Hypothesis Testing



Four-cluster Likelihood Mapping - Specific Hypothesis Testing

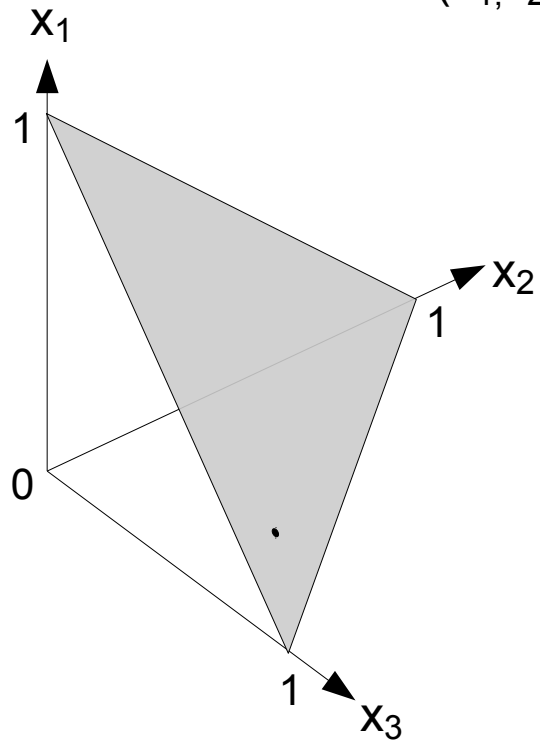


Four-cluster Likelihood Mapping - Specific Hypothesis Testing

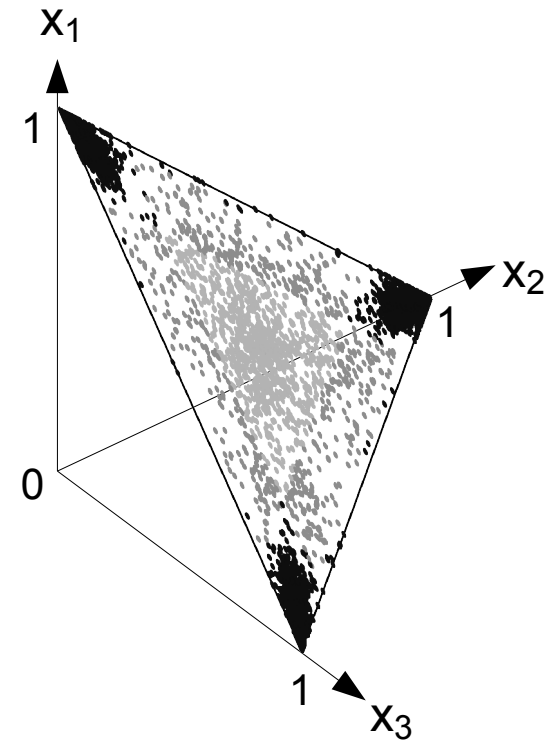
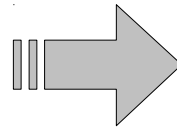
$$S(x_1, x_2, x_3) = (0.1, 0.1, 0.8)$$

$$S(x_1, x_2, x_3) = (0.3, 0.3, 0.3)$$

$$S(x_1, x_2, x_3) = (0.7, 0.1, 0.2)$$



⋮





Four-cluster Likelihood Mapping - Specific Hypothesis Testing

Example: Basal Hexapods

G1: Protura, G2: Diplura, G3: Collembola, G4: Ectognatha *partim*

Possible unambiguous topologies:

- T1: G1G2 | G3G4 = Protura+Diplura
- T2: G1G3 | G2G4 = Protura+Collembola
- T3: G1G4 | G2G3 = Diplura+Collembola

Possible ambiguous topologies:

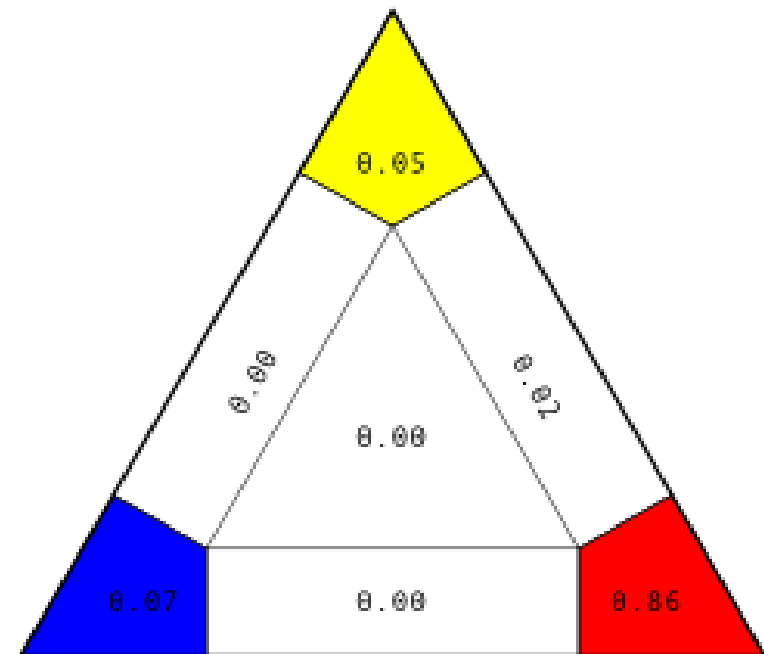
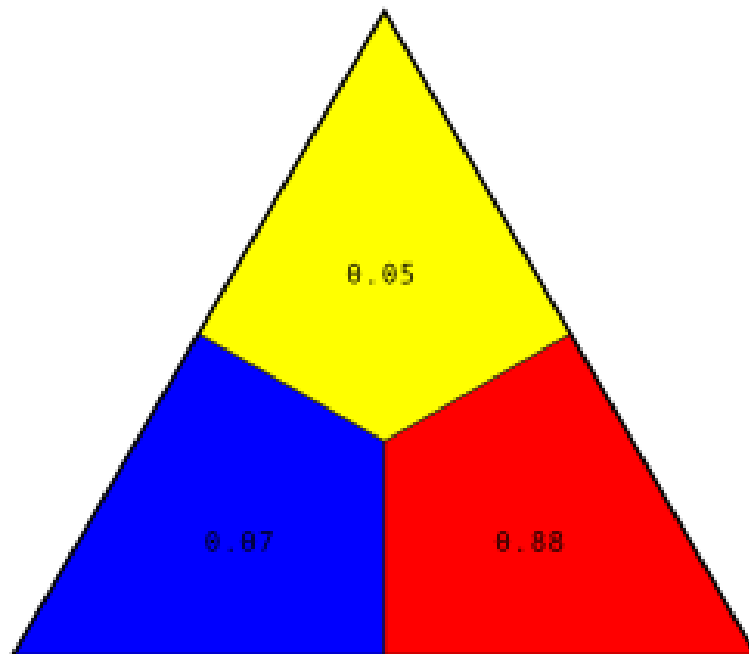
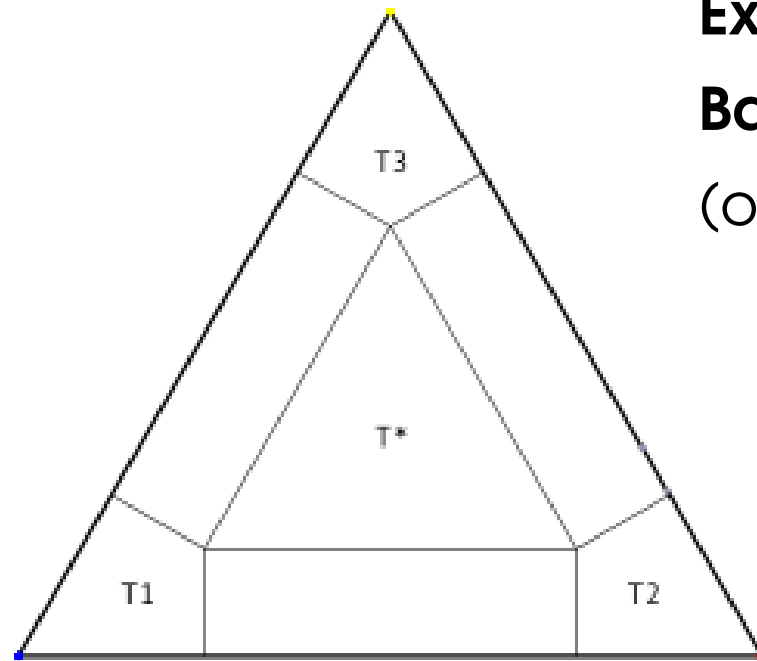
- T1T2: G1G2 | G3G4 or G1G3 | G2G4
- T1T3: G1G2 | G3G4 or G1G4 | G2G3
- T2T3: G1G3 | G2G4 or G1G4 | G2G3
- T*: G1G2 | G3G4 or G1G3 | G2G4 or G1G4 | G2G3

Number of all possible quartets: 120

LM_simplex

number of quartets: 120

Example:
Basal Hexapods
(original)



A close-up photograph of a fly's head and antenna, showing the compound eye and the segmented structure of the antenna. The fly is positioned in the upper half of the frame, with its head and antenna extending towards the right. The background is a soft, out-of-focus green, suggesting an outdoor environment.

Four-cluster Likelihood Mapping - Specific Hypothesis Testing

**Evaluate result with the help of permutations:
compare original data with**

- permutation I

→ permute non-ambiguous amino acid characters, leaving the distribution of missing characters untouched

- permutation II

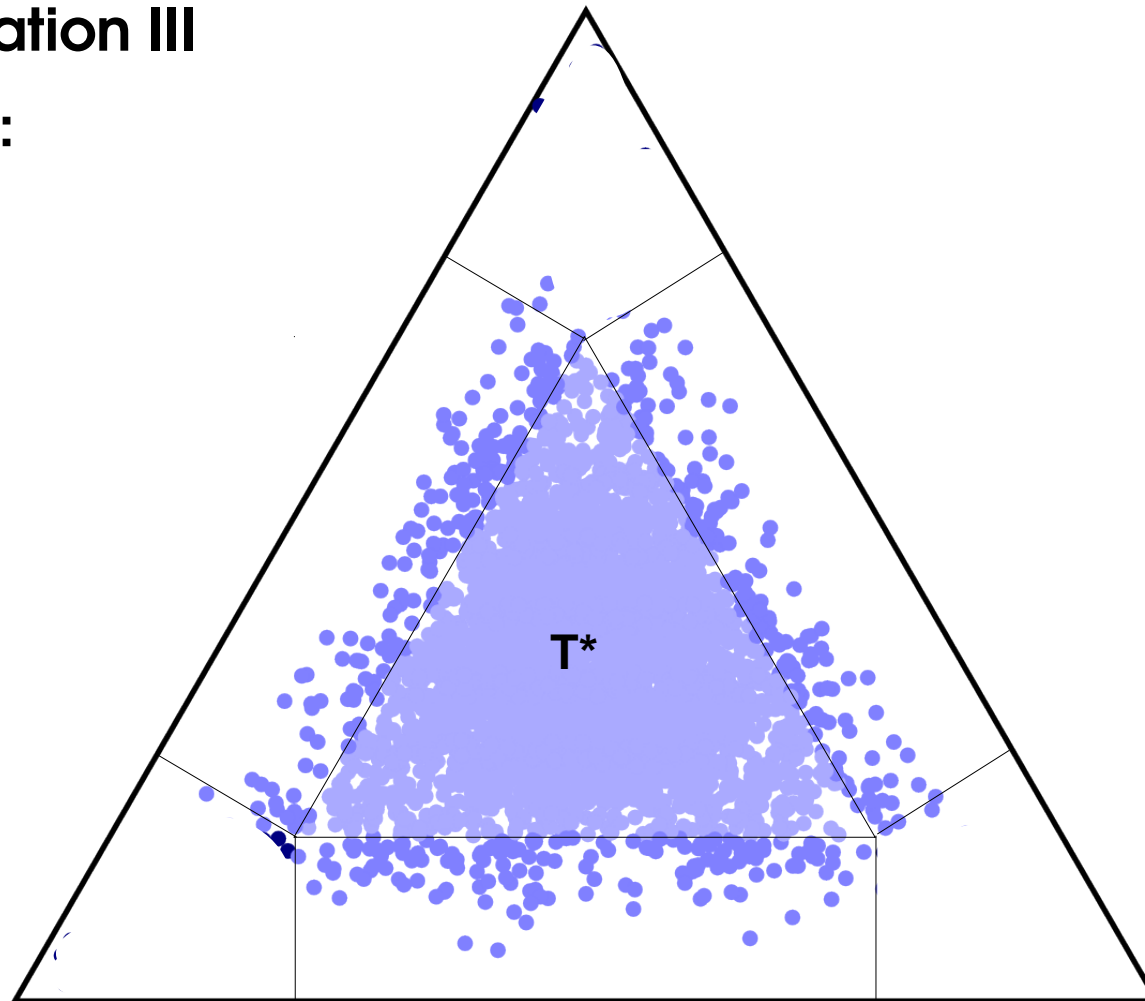
→ see permutation I but, use the WAG matrix to generate partitions that match stationary, homogeneous and reversible (SHR) conditions

- permutation III

→ see permutation II but, additionally with randomised missing data to obtain a matrix that matches SHH conditions and with randomly distributed missing data.

Four-cluster Likelihood Mapping - Specific Hypothesis Testing

for permutation III
we expect:



Four-cluster Likelihood Mapping - Specific Hypothesis Testing

Supermatrix C: Topology	original (support in %):	permutation I (support in %):	permutation II (support in %):	permutation III (support in %):
T1	7	9	47	12
T2	86	23	7	27
T3	5	54	4	5
T1T2	0	4	12	12
T1T3	0	3	15	3
T2T3	2	5	4	23
T*	0	1	12	21

T2: G1G3 | G2G4 = Ellipura

Acknowledgements

- David Yeates, ANIC, CES
- the 1KITE consortium, especially Bernhard Misof (zmb)
- the Beijing Genomics Institute (BGI)
- Lars Jermiin & group, Bioinformatics & Phylogenomics, CES
- Rob Lanfear, Brett Calcott, ANU
- Barbara Holland, Jeremy Sumner
- *all further collaborators, collectors, identifiers, etc that are continuously in this projects...*



Thank you!

www.1kite.org

