Advancing Collection Genomics

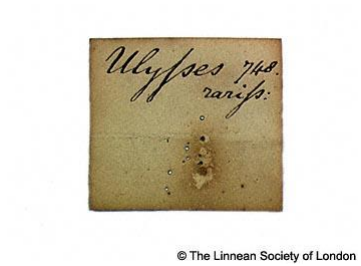
Andreas Zwick , Beth Mantle

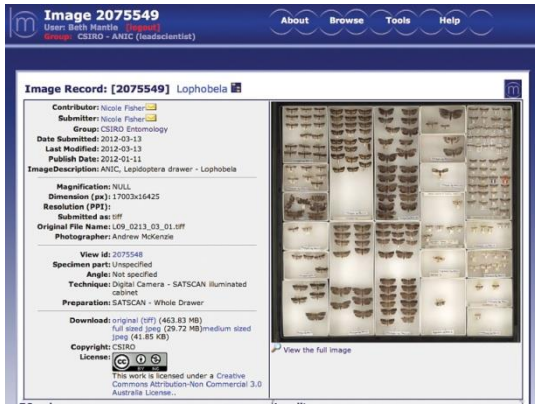
14 April 2015

Digitising the Blueprint of Life

Collection genomics – the ultimate way of digitising?

- Digitising



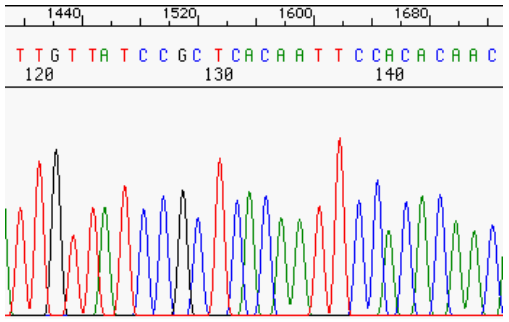


- Genome sequencing = Digitising the entire blueprint of life

Chemical structure

A C T G

ASCII file




```
>Homo_sapiens.20150414
CAATACAGGATACATGGAAA
TATTAGACCATGGGACCCTA
CATGGACTACCATATCCATG
GACTATAGGCATCAGAAAAT
GGGATATCCATAGATCATAA
CTCTCTAATGATAGCAGTCG
```

Digitising the Blueprint of Life

Collection genomics – the ultimate way of digitising?

- **BUT**: Whole genome alone cannot account for the phenotype
  - Influence of environment
  - Gene expression patterns
  - DNA methylation
  - Small interfering RNA (siRNA): Gene silencing

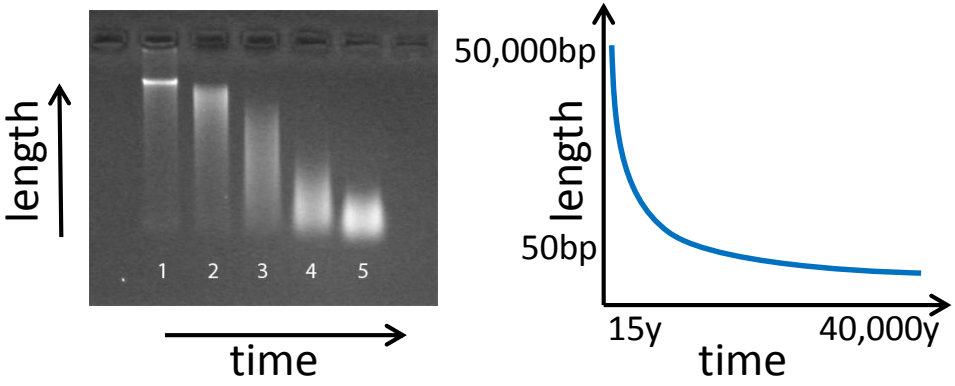
Digitising the Blueprint of Life

Collection genomics – the ultimate way of digitising?

- Most collection specimens still contain their entire genome
  - DNA degrades over time

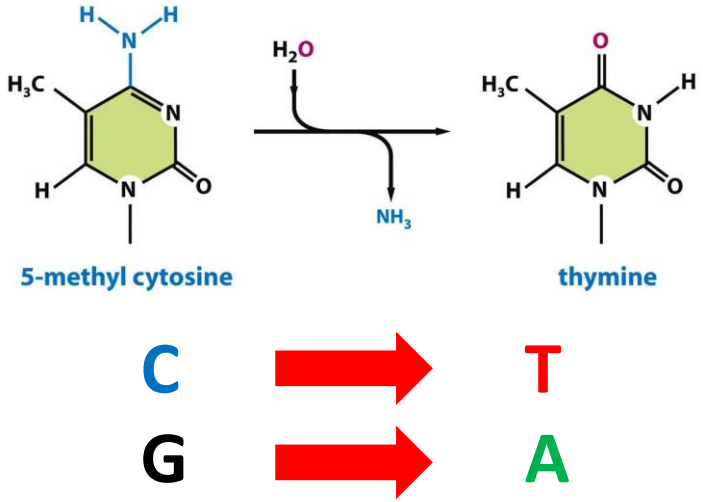
RAPIDLY

Breaks in DNA strand:
Short fragments



SLOWLY

Deamination of bases:
Sequencing errors



Digitising the Blueprint of Life

Entire genomes from ancient collection specimens

CBA-funded

Example: *Helicoverpa armigera* and *H. prepodes*

- *H. armigera*, the most damaging agricultural pest in Australia
  - Recorded from more than 200 host plants
  - Highly abundant in Australia
  - genome sequenced by CSIRO and collaborators
- *H. prepodes*, a very rarely collected native species
  - No fresh samples available for genome sequencing
  - 3% CO1 divergence to *H. armigera* and all other *Helicoverpa* spp.
- Pilot study to sequence genomes from collection specimens
  - 50-100 years old: DNA strongly fragmented (25-50bp)
  - About 90% success mapping reads against reference genome for both spp.

H. prepodes

Whole genome (re-)sequencing is possible,
but tedious and requires similar reference genome

Digitising the Blueprint of Life

What can we do with it?

- Stabilise nomenclature

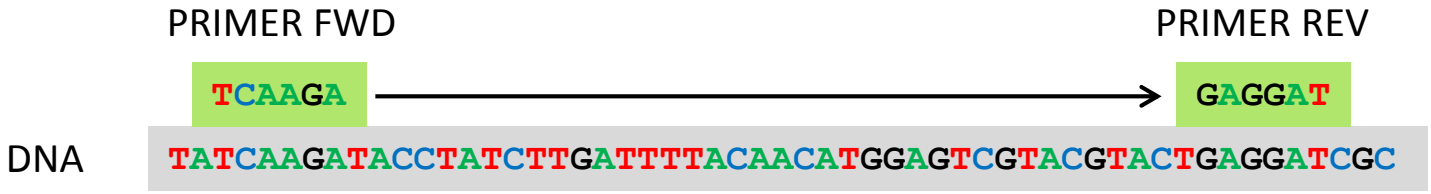
  e.g., *Drosophila* Fallén, 1823 is a junior synonym of *Oinopota* Kirby & Spence, 1815

- Sequence type specimens
- Identify large quantities of samples
- Build phylogenies : Evolutionary, predictive frameworks
- Assess genetic diversity: Conservation
- Screen for resistance genes and trace their development over time
- Screen digitised geno- and phenotypes for patterns

Rapid digitisation of partial genomes

The traditional approach – PCR for DNA barcoding

- Selective target enrichment through PCR amplification
  - A pair of specific primers binds to DNA
  - Fragment between primers gets amplified and sequenced
  - Requires template DNA fragments that are longer than target
    - Not suitable for highly degraded DNA
    - Typically works only with “recent” specimens <15 years old



- If all goes well: $15 / sample and gene

Rapid digitisation of partial genomes

First large-scale project in ANIC – “The Barcode Blitz”

- Paul Hebert and team (Biodiversity Institute of Ontario, CANADA)
- Two visits, 5 weeks each, 5 people each: 28,000 samples
- Processing pipeline

Labelling

Databasing and geocoding

Imaging

Sampling (leg)

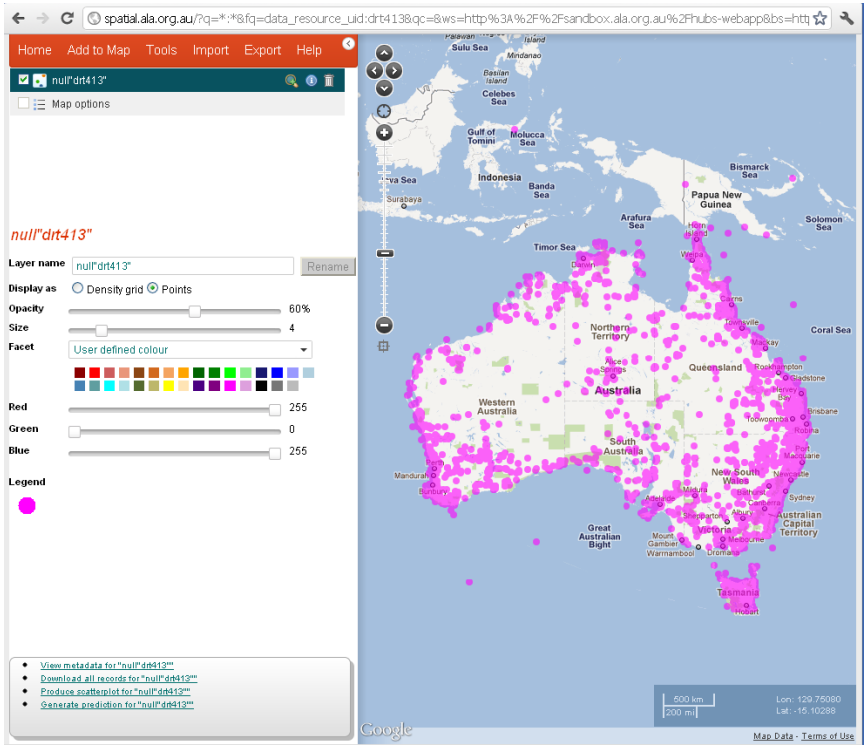
- Sequencing in Canada

Rapid digitisation of partial genomes

First large-scale project in ANIC – "The Barcode Blitz"

- 28,000 samples of 8,000 named species in 10 weeks
  - 80% of named fauna
  - Continent-wide coverage
  - CO1 reference library
  - Discovery of cryptic species, e.g., *Plutella australiana*

- Some well-known limitations
  - Single gene !!!
  - Mitochondrial, maternally inherited, *Wolbachia* influence, NUMTs, ...



Sampling 28,000 specimens / 8,000 species in a collection during a short period (10 weeks) is feasible!

Rapid digitisation of partial genomes

Current large-scale project in ANIC – transcriptomes of insects

- Transcriptome = RNA transcripts of all “active” genes

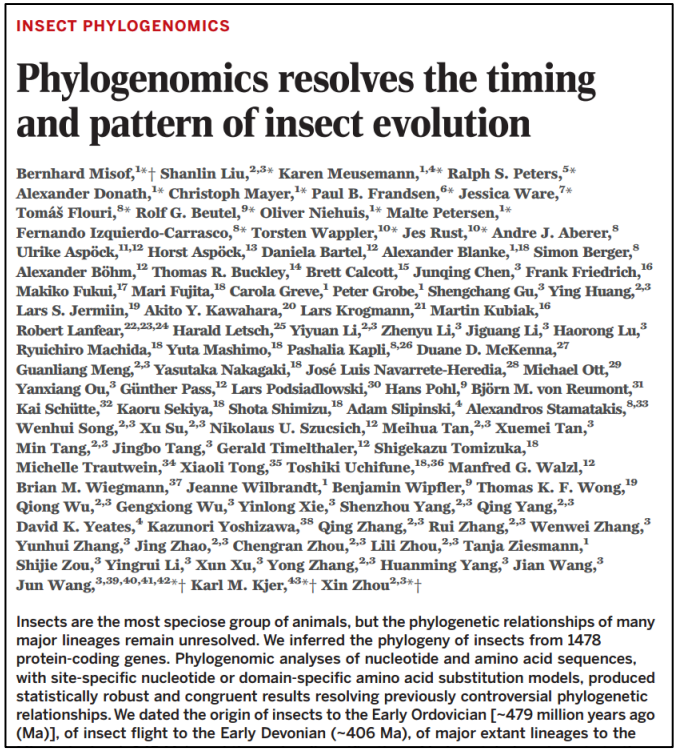
- Sequencing of 600 transcriptomes from fresh Australian insects
  - moths (Lepidoptera)
  - beetles (Coleoptera)
  - flies (Diptera)

- Lepidoptera
  - 10,000 named, and as many unnamed, species in Australia
  - Use morphology-based classification and CO1 barcodes for guidance: Sample as divergent lineages as possible

Rapid digitisation of partial genomes

Current large-scale project in ANIC – transcriptomes of insects

- Huge genomic resource for higher phylogenomics
  - Family- and superfamily level
  - 600 samples x ~15,000 genes, based on 1,800,000,000bp
- 1KITE project unravelled evolution of insects (Misof *et al*. 2014. *Science*)



INSECT PHYLOGENOMICS

Phylogenomics resolves the timing and pattern of insect evolution

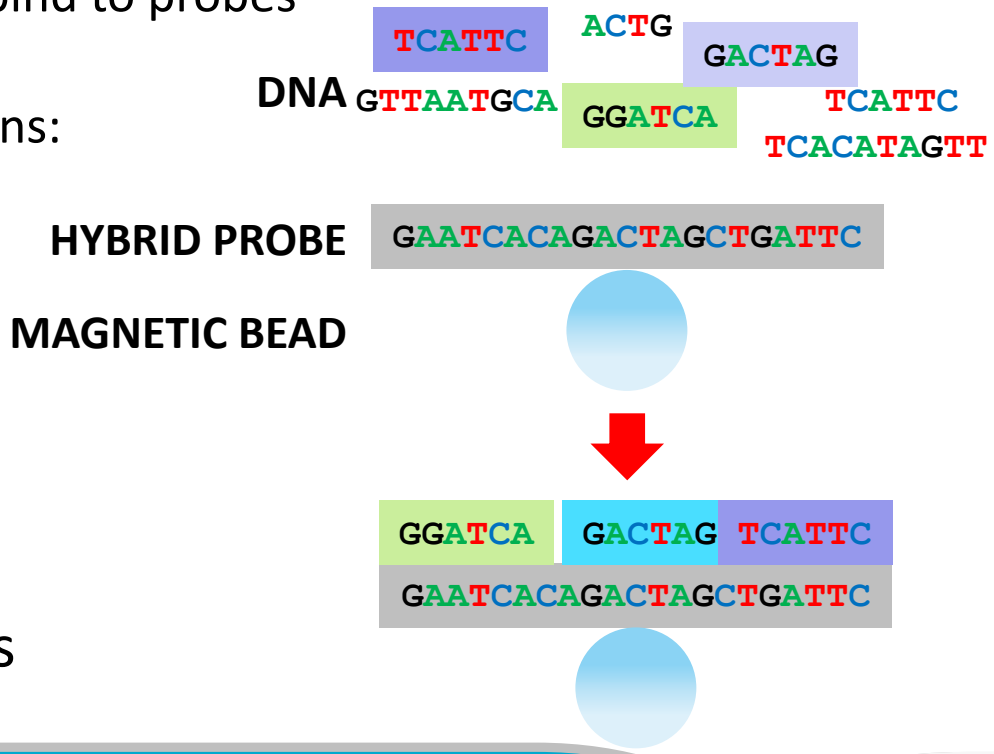
Bernhard Misof,[1*†] Shanlin Liu,[2,3*] Karen Meusemann,[1,4*] Ralph S. Peters,[5*] Alexander Donath,[1*] Christoph Mayer,[1*] Paul B. Frandsen,[6*] Jessica Ware,[7*] Tomáš Flouri,[8*] Rolf G. Beutel,[9*] Oliver Niehuis,[1*] Malte Petersen,[1*] Fernando Izquierdo-Carrasco,[8*] Torsten Wappler,[10*] Jes Rust,[10*] Andre J. Aberer,[8] Ulrike Aspöck,[11,12] Horst Aspöck,[13] Daniela Bartel,[12] Alexander Blanke,[1,18] Simon Berger,[8] Alexander Böhm,[12] Thomas R. Buckley,[14] Brett Calcott,[15] Junqing Chen,[3] Frank Friedrich,[16] Makiko Fukui,[17] Mari Fujita,[18] Carola Greve,[1] Peter Grobe,[1] Shengchang Gu,[3] Ying Huang,[2,3] Lars S. Jermiin,[19] Akito Y. Kawahara,[20] Lars Krogmann,[21] Martin Kubiak,[16] Robert Lanfear,[22,23,24] Harald Letsch,[25] Yiyuan Li,[2,3] Zhenyu Li,[3] Jiguang Li,[3] Haorong Lu,[3] Ryuichiro Machida,[18] Yuta Mashimo,[18] Pashalia Kapli,[8,26] Duane D. McKenna,[27] Guanliang Meng,[2,3] Yasutaka Nakagaki,[18] José Luis Navarrete-Heredia,[28] Michael Ott,[29] Yanxiang Ou,[3] Günther Pass,[12] Lars Podsiadlowski,[30] Hans Pohl,[9] Björn M. von Reumont,[31] Kai Schütte,[32] Kaoru Sekiya,[18] Shota Shimizu,[18] Adam Slipinski,[4] Alexandros Stamatakis,[8,33] Wenhui Song,[2,3] Xu Su,[2,3] Nikolaus U. Szucsich,[12] Meihua Tan,[2,3] Xuemei Tan,[3] Min Tang,[2,3] Jingbo Tang,[3] Gerald Timelthaler,[12] Shigekazu Tomizuka,[18] Michelle Trautwein,[34] Xiaoli Tong,[35] Toshiki Uchifune,[18,36] Manfred G. Walzl,[12] Brian M. Wiegmann,[37] Jeanne Wilbrandt,[1] Benjamin Wipfler,[9] Thomas K. F. Wong,[19] Qiong Wu,[2,3] Gengxiong Wu,[3] Yinlong Xie,[3] Shenzhou Yang,[2,3] Qing Yang,[2,3] David K. Yeates,[4] Kazunori Yoshizawa,[38] Qing Zhang,[2,3] Rui Zhang,[2,3] Wenwei Zhang,[3] Yunhui Zhang,[3] Jing Zhao,[2,3] Chengran Zhou,[2,3] Lili Zhou,[2,3] Tanja Ziesmann,[1] Shijie Zou,[3] Yingrui Li,[3] Xun Xu,[3] Yong Zhang,[2,3] Huanming Yang,[3] Jian Wang,[3] Jun Wang,[3,39,40,41,42*†] Karl M. Kjer,[43*†] Xin Zhou[2,3*†]

Insects are the most speciose group of animals, but the phylogenetic relationships of many major lineages remain unresolved. We inferred the phylogeny of insects from 1478 protein-coding genes. Phylogenomic analyses of nucleotide and amino acid sequences, with site-specific nucleotide or domain-specific amino acid substitution models, produced statistically robust and congruent results resolving previously controversial phylogenetic relationships. We dated the origin of insects to the Early Ordovician [~479 million years ago (Ma)], of insect flight to the Early Devonian (~406 Ma), of major extant lineages to the

Rapid digitisation of partial genomes

Next large-scale project in ANIC – multi-gene next-gen barcoding

- Target enrichment for 700 genes with hybridisation probes
  - Transcriptomes are used to identify gene targets
  - Transcriptomes provide templates to synthesize probes
  - Matching sample DNA fragments bind to probes (10-15% divergence OK)
  - Suitable for old collection specimens:
    - No specific primers
    - No need for large fragments
    - No need for clean samples

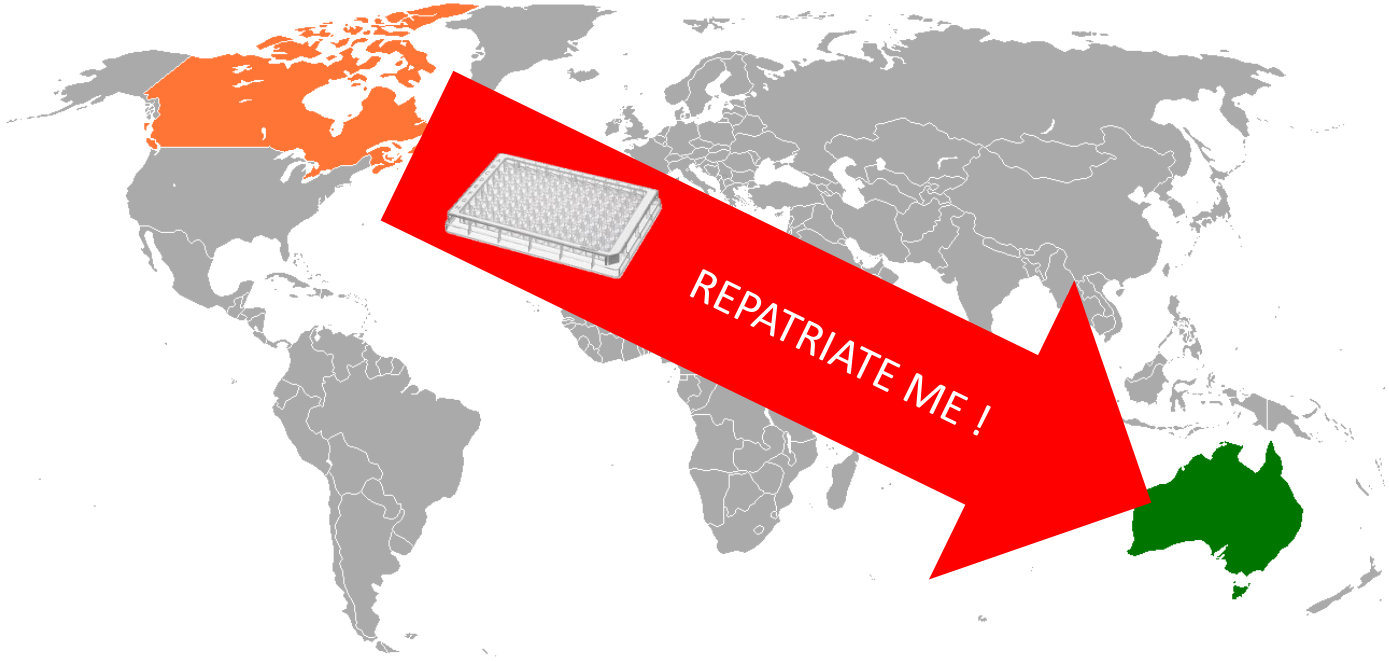


- < $100 / sample and 700 genes

Rapid digitisation of partial genomes

Next large-scale project in ANIC – multi-gene next-gen barcoding

- Ideal first test case: Extracts from BOLD barcoding campaign
  - Direct comparison of single- and multi-gene barcoding
  - Labour-intense physical sampling and documenting already done
  - No further physical damage to 28,000 samples



Advancing Collection Genomics

A straight-forward road for the ANIC

Thank you

Australian National Insect Collection
Andreas Zwick

t +61 2 6246 4289
e andreas.zwick@csiro.au
w www.csiro.au/en/Research/Collections/ANIC

National Research Collections Australia
Beth Mantle

t +61 2 6246 4281
E beth.mantle@csiro.au
w www.csiro.au/en/Research/Collections