ON THE ROAD TO FULL TAXON SAMPLING FOR PHYLOGENETIC WORK ON NYMPHALIDAE: LESSONS FROM A MASSIVE DATASET

Niklas Wahlberg
CHAOS IN NATURE?

Bewildering diversity
Unevenly distributed
Different kinds of interactions
Different kinds of reactions

HOW DO WE UNDERSTAND THIS?

Despite bewildering diversity, there are distinct patterns
Uneven distributions appear to be correlated with external factors
Interactions can be classed into general groups
Reactions are predictable?
EVERYTHING HAS A HISTORY

We can understand the present by looking at the past
We hope to predict the future by understanding the present
How can we get at the past?

PHYLOGENIES ARE NEEDED

Depict the evolutionary relationships of taxa we are interested in
Cannot be observed, only inferred
Getting a robust estimate is important
PHYLOGENIES ARE NEEDED

Depict the evolutionary relationships of taxa we are interested in
Cannot be observed, only inferred
Getting a robust estimate is important

THE VERSATILITY OF DNA SEQUENCES

DNA sequences are a record of the evolutionary history of an individual
- Can be used to reconstruct the phylogenetic relationships of species groups
- Can be used to infer times of divergence
THE POSTGENOMIC ERA IS HERE

Generating genome data is easy, quick and getting cheaper
The first transcriptome based analyses have already been published
Huge amount of data
Do we need so much data, or will less be enough?

RESOLVING A REIGN OF CHAOS

The butterfly family Nymphalidae
About 6000 species
Distributed globally
Highest species richness in South America
CLASSIFICATION
Classification of the about 6000 species has been a source of contention
Sometimes split into 9 different families!
Some taxa have remained relatively stable (e.g. Brassolini, Libytheinae, Melitaeini)
Other taxa have changed a lot (e.g. Limenitidinae, Nymphalinae, Satyrinae)

PHYLOGENY?
Higher level systematics was unknown prior to molecular revolution!

```
  Libytheinae
 /       \
Danainae  Limentidinae
 /     \   /     \
Apaturinae Heliconiinae
 /                 /     \
Nymphalinae Morphinae Satyrinae
               /     \    
            Calinaginae Charaxinae
```

PHYLOGENY BEGINS TO UNRAVEL

3 studies pioneered the way:
  • Brower 2000: based on wingless gene
  • Wahlberg et al 2003: based on three genes
  • Freitas & Brown 2004: based on morphology

These studies have been followed by many studies culminating in:
  • Wahlberg et al 2009: based on ten genes and morphology
Wahlberg et al. 2009: PRS 276


satyrine clade

Figure 1: Compiled view of species of R. Only taxonomic names from the Revision section. Rheten is now re-classified for seeing the species of satyrine clade and species of satyrine clade.
BUT DOUBT REMAINS...

Just how robust are these conclusions?
Are 10 genes in fact enough?
What about taxon sampling, only 6% of species sampled?
Unpublished mitochondrial genome analyses seem to be saying something very different
TIME TO SEE WHERE WE ARE WITH TAXON SAMPLING

More than 2000 species of Nymphalidae have been sequenced for between 3 and 11 genes

VoSeq database, managing sequences from 11 genes

Peña & Malm 2012: PlosOne 7: e39071
2056 species of Nymphalidae

TOO LARGE AND TOO CUMBERSOME

Still calculating bootstrap values
Not possible to analyze in a Bayesian framework
Can this information be retrieved with smaller taxon sampling?
Minimizing taxon sampling

One per subfamily:
12 species

One per tribe:
45 species
200 taxa

One per genus: 454 species
Two per genus (when available): 661 species

Diversity represented: 916 species
THE EFFECTS OF TAXON SAMPLING

Clearly more is better, but how much more? A trade-off between sample size and computing time
For deep phylogeny of Nymphalidae, aiming for one species per genus appears to be optimal
Analyzing 454 taxa today is much quicker than 5 years ago!
WHERE ARE WE WITH NYMPHALIDAE PHYLOGENY?

Subfamilies are now completely clear, well-supported monophyletic groups
Relationships of subfamilies appear to be stable

STILL MANY OPEN QUESTIONS

Relationships within subfamilies, e.g. of tribes?
- 916 taxon dataset ideal for this
Effects of taxon sampling on times of divergence?
What about more genes? Transcriptomes?
Mitochondrial genomes?
THANKS TO

• NSG
  • Carlos Peña, Tobias Malm, Reza Zahir, Julien Leneveu, Ullasa Kodandaramaiah, Pavel Matos

• Lepidoptera phylogeny
  • Lauri Kaila, Marko Mutanen, Maria Heikkilä, Jadranka Rota

• Phylogenomics
  • Chris Wheat

• Money
  • Academy of Finland, Vetenskapsrådet, Kone Foundation