

Metagenome skimming of species-rich lineages



Imperial College
London



Benjamin Linard
b.linard@nhm.ac.uk

Follow me on

ResearchGate

Who is this guy ?

Thesis in bioinformatics: IGBMC Strasbourg, bioinfo lab of Julie Thompson, Olivier Poch



- Java software development
- Orthology/paralogy inference via algorithmics
- Online databases development
- Cluster and network-based visualization of protein-based evolutionary histories

- comparative genomics of mammals proteomes

Post-doc in environmental genomics: NHM London, phylogenetics lab of Alfried Vogler



- NGS data processing
- metagenomics
- large-scale taxonomic assignments
- mitochondria analyses and annotations
- insect genomics and comparative genomics

- analyses of species-rich communities
(insects and bacterial symbionts)

Thesis work: orthology and paralogy

Interoperability

- Compatible with any SQL engine
- Platform independent
- OrthoXML specifications
- Pipeline friendly



Bioinformatics

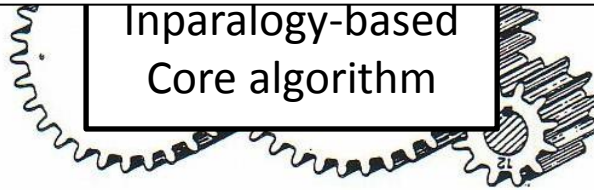
ABOUT THIS JOURNAL CONTACT THIS JOURNAL SUBSCRIPTIONS CURRENT

Oxford Journals > Science & Mathematics > Bioinformatics > Volume 31, Issue 3 > Pp. 447-448

OrthoInspector 2.0: Software and database updates

Benjamin Linard^{1,2}, Alexis Allot¹, Raphaël Schneider¹, Can Morel¹, Raymond Ripp¹, Marc Bigler¹, Julie D. Thompson¹, Olivier Poch¹ and Odile Lecomte^{1,*}

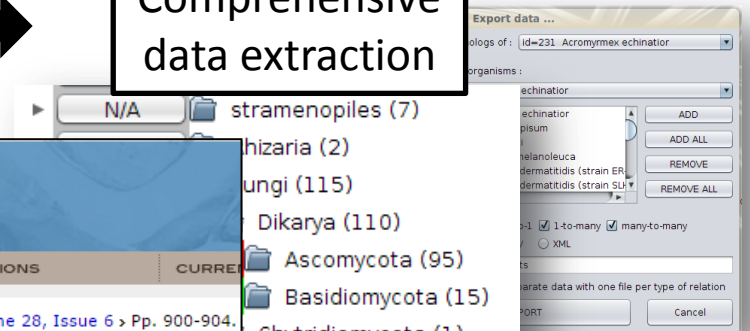
Inparalogy-based Core algorithm



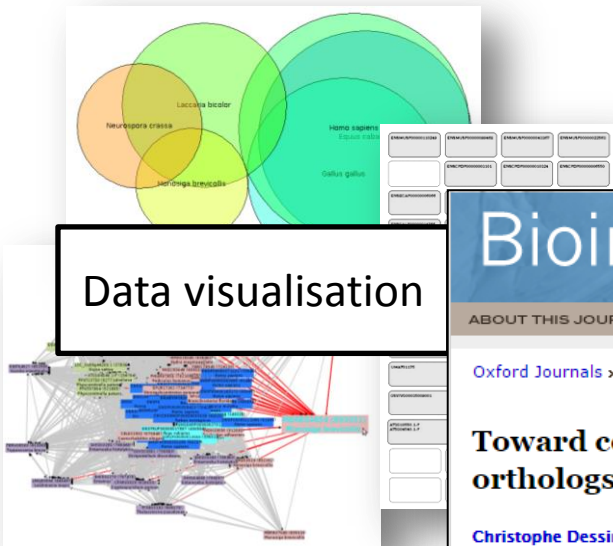
Online database

lbgi.fr/orthoinspector/

Comprehensive data extraction



Data visualisation



Bioinformatics

ABOUT THIS JOURNAL CONTACT THIS JOURNAL SUBSCRIPTIONS CURRENT

Oxford Journals > Science & Mathematics > Bioinformatics > Volume 28, Issue 6 > Pp. 900-904.

Toward community standards in the quest for orthologs

Christophe Dessimoz^{1,*}, Toni Gabaldón², David S. Roos³, Erik L. L. Sonnhammer⁴, Javier Herrero⁵ and the Quest for Orthologs Consortium[†]

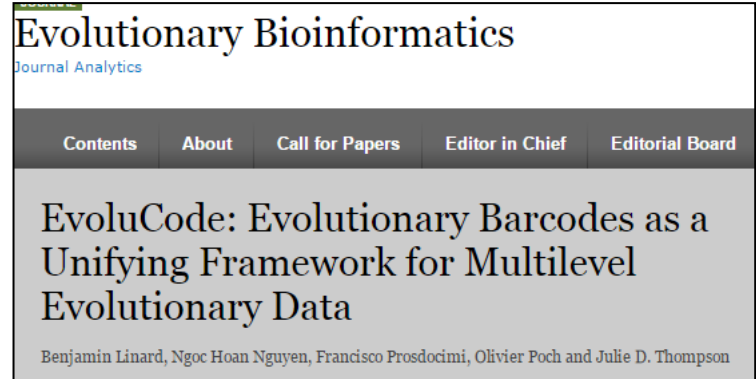
+ Author Affiliations

Thesis work: visualization tools

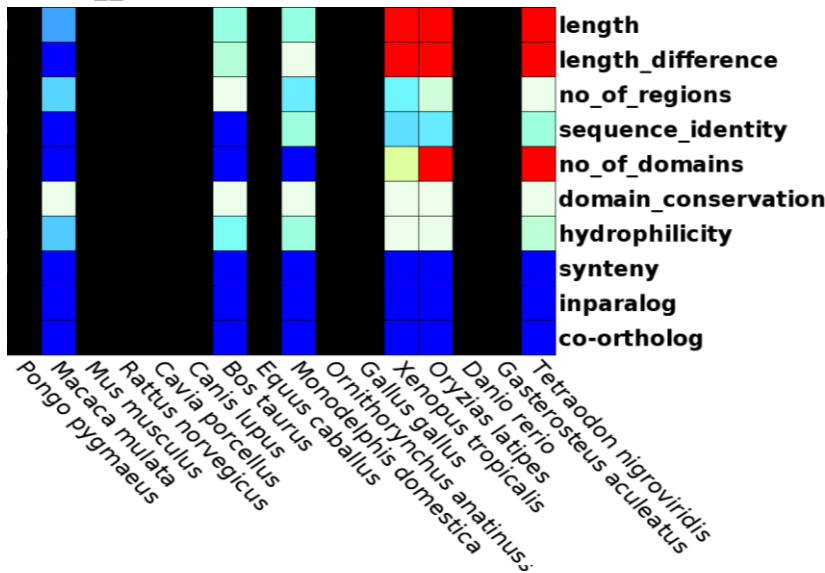
EvoluCodes : Evolutionary Barcodes



1 gene
1 evolutionary history
1 barcode



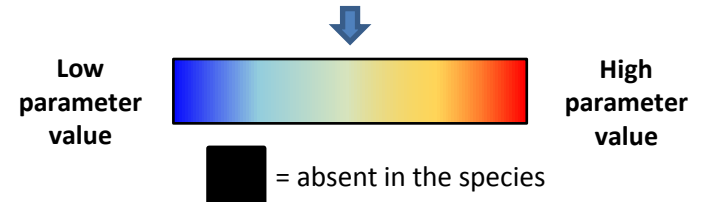
HBV 16 19 22 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000



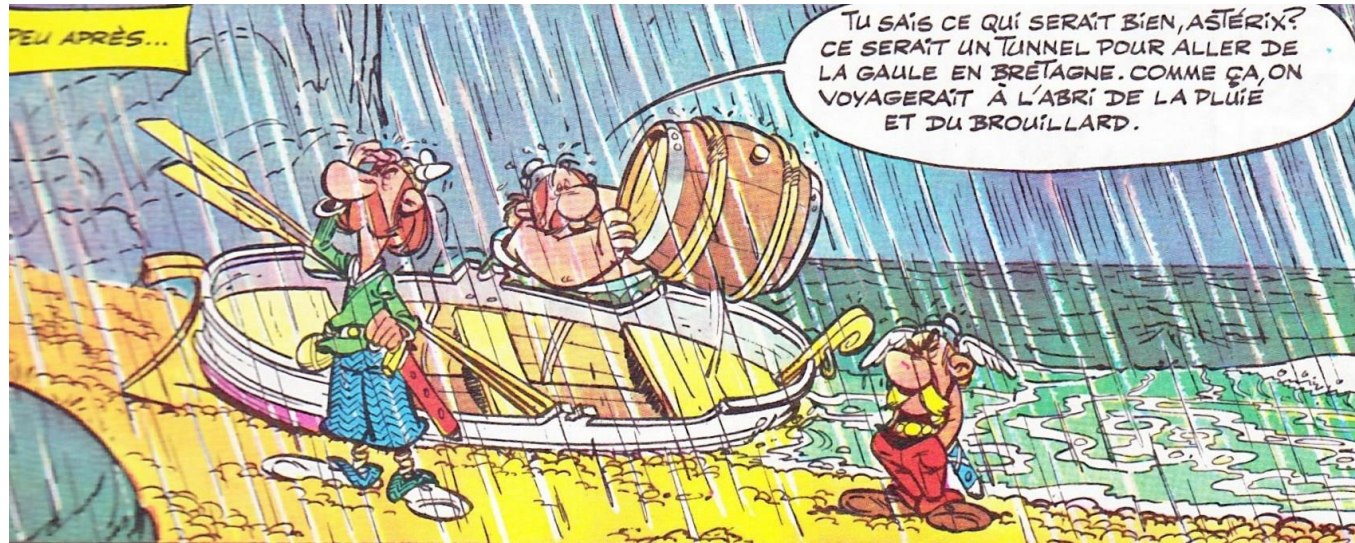
- Integrates multi-scale data
- Describes the variation of a gene
- A framework for knowledge extraction
- Facilitates visualisation of biological messages

Variable repartition in vertebrates, viral DNA integration

Generally observed value



Moving to the Natural History Museum



(publié en 1966, l'eurotunnel n'existait pas encore !)

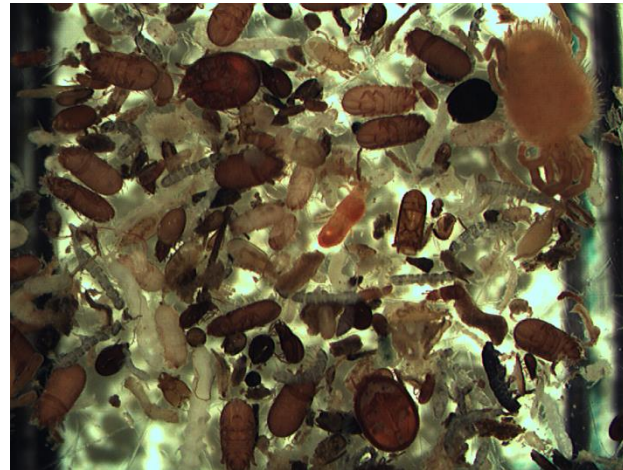
From algorithmics, Java software development
and comparative genomics ...

...to the development of new methodologies to generate
de-novo RAW data
(and understand them via comparative genomics)

Context

Arthropods in current ecosystems

- 1,300,000 arthropods species described (80% of all described animals)
estimates suggest 2 to 20 million species (Basset et al. 2012; Zhang 2011)
- Virtually present in all ecosystems. A tremendous source of biodiversity and genomic diversity!



Credits: C. Andujar, P. Arribas

- Arthropod diversity is modulated by : habitat health, pollution, climate change.
- They are an “easy to reach” bioindicator

Context

Arthropod genomics

- Only dozens of insect complete genome and transcriptomes, most of them Diptera or relevant to human health / pest control
- only two complete Coleopteran complete nuclear genomes in 2015 :
Tribolium castaneum (Tenebrionoidea) and *Dendroctonus ponderosae* (Curculionoidea)
(Friedrich & Muqim 2003; Keeling et al. 2013).

Understanding the black box between

“molecular community” and “ecological community” (Huttenhower & Hofmann 2010)

Big Initiatives...
But slow process



- **Arthropod biodiversity studies generally focus on specific loci, which are targeted mostly through Metabarcoding approaches.**

Context

What about metagenomics approaches ?

**Genome skimming:
Non-targeted and non-selective**

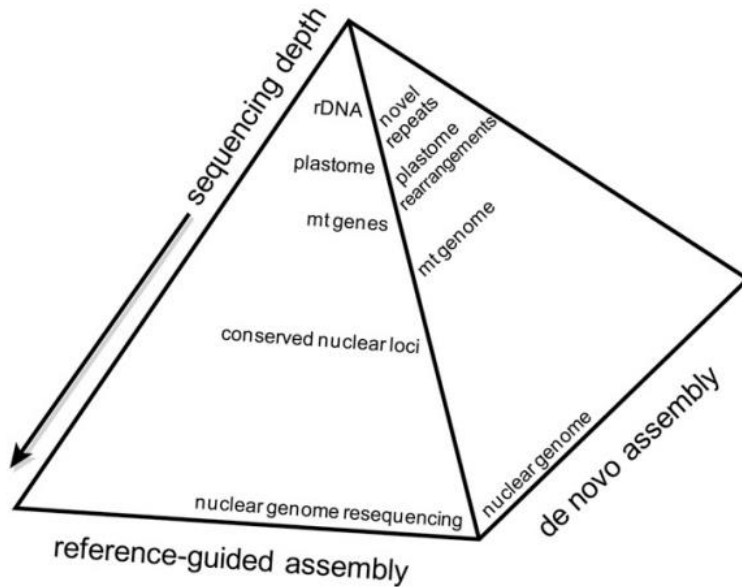


Fig. 1. The genomic iceberg: the relationship between genomic targets, the sequencing depth required to obtain them, and the most appropriate method of sequence assembly.

(Straub et al., 2012)

MOLECULAR ECOLOGY RESOURCES

Resource Article



Australian
National
University

Centre for Biodiversity Analysis



menu

Home » Research » Projects » Genome skimming with degraded DNA from herbarium specimens

Genome skimming with degraded DNA from herbarium specimens



Members

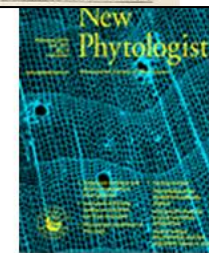
Researcher

- Assoc. Prof. Adrienne Nicotra
- Dr Alexander Schmic Lebuhn

Article first published online: 18 NOV 2013

DOI: 10.1111/nph.12560

© 2013 The Authors. New Phytologist © 2013
New Phytologist Trust



New Phytologist

Volume 201, Issue 3, p
1021–1030, February 2013

Context

Genome skimming:

→ shallow sequencing of direct DNA extractions

**Non-targeted
and non-selective**

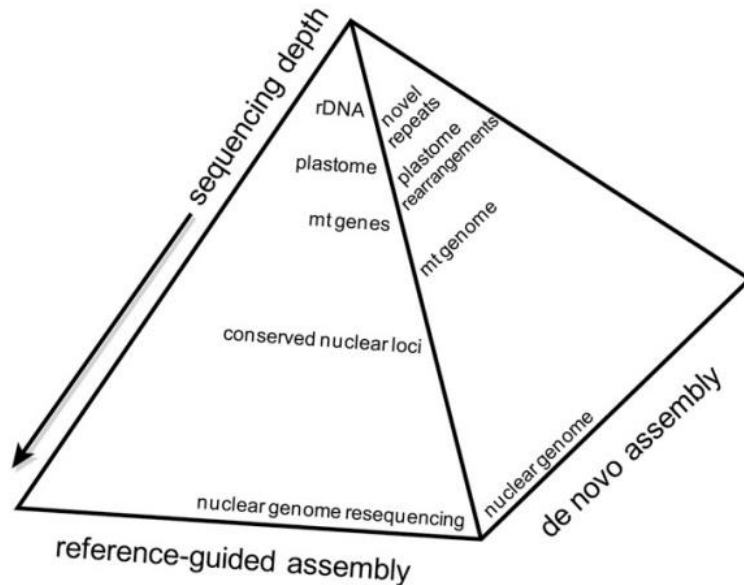


Fig. 1. The genomic iceberg: the relationship between genomic targets, the sequencing depth required to obtain them, and the most appropriate method of sequence assembly.

(Straub et al., 2012)

- **De-novo assembly of the reads**
- Only the most abundant DNA motifs are assembled.
- Organelles are the first obvious outcome (many genome copies per cell)
- Genomic repeats were recently proposed as another source of phylogenetic signal

Systematic Biology

ABOUT THIS JOURNAL CONTACT THIS JOURNAL SUBSCRIPTIONS

Oxford Journals > Life Sciences > Systematic Biology > Volume 64, Issue 1 > Pp. 112

Genomic Repeat Abundances Contain Phylogenetic Signal

Steven Dodsworth^{1,2}, Mark W. Chase^{2,3}, Laura J. Kelly^{1,2}, Ilia J. Leitch², Jiří Macas⁴, Petr Novák⁴, Mathieu Piednoël⁵, Hanna Weiss-Schneeweiss⁶ and Andrew R. Leitch^{1,*}

[+](#) Author Affiliations

Mitochondrial metagenomics

(or mito-metagenomics or mitogenomics...)

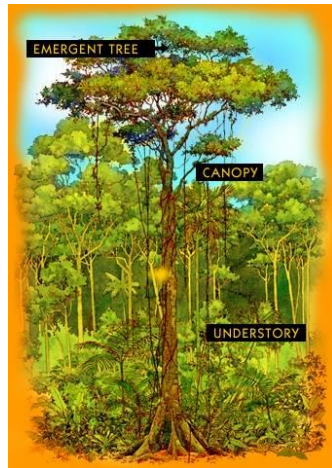
Mitochondria DNA is abundant in animal cells, untargeted NGS of arthropods samples around 1% of mitochondrial reads.



Pool of 50 to 300 species
(hundreds of specimens)

→ goal: complete mitochondrial genomes
a proxy to biodiversity and phylogenetic signal

The NHM Biodiversity Initiative: Reaching rapidly patterns of arthropod biodiversity



n traps per site:
(soil, canopy,
Ground...)



N plots

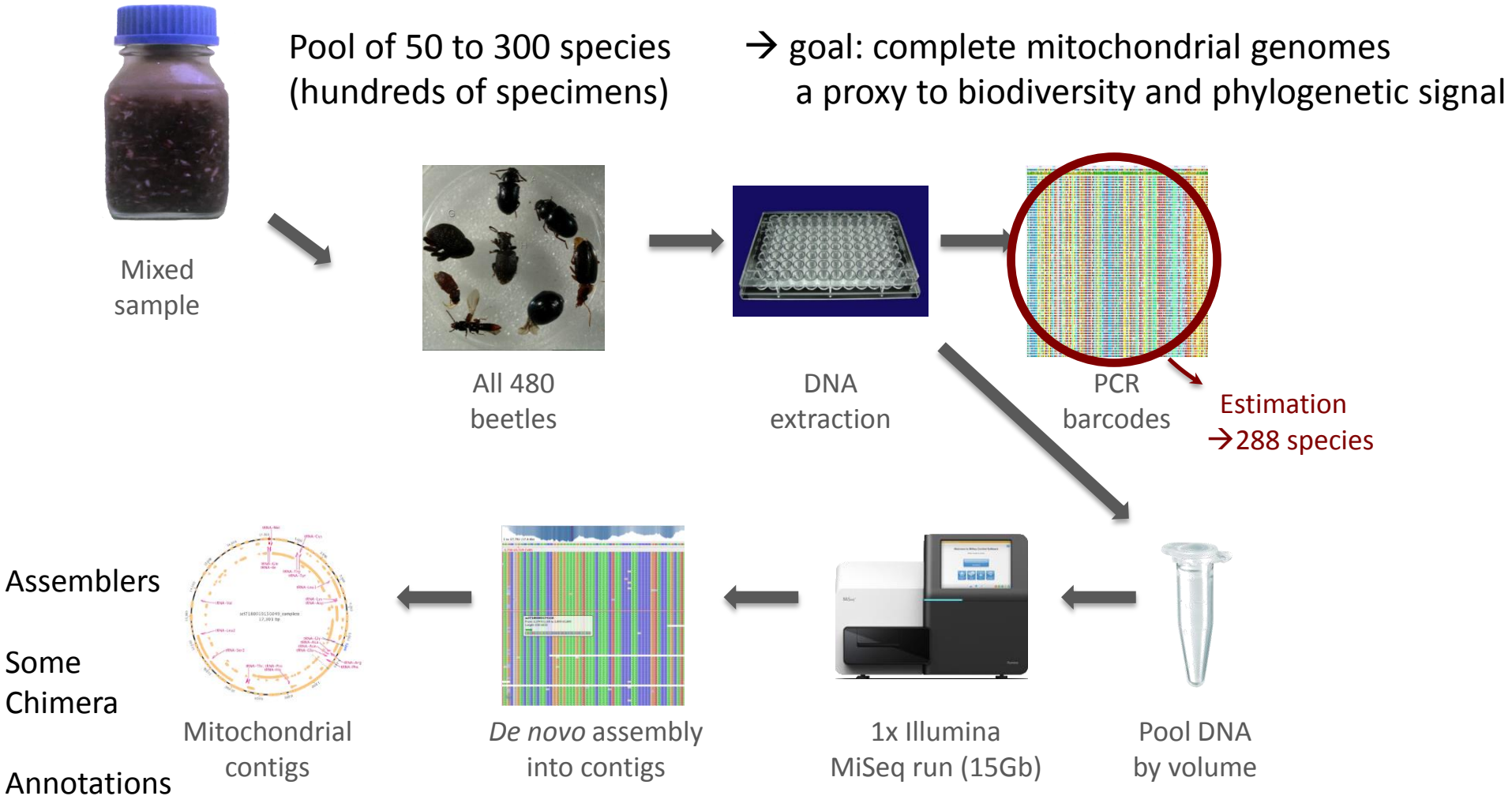


Many soup samples...



Mitochondrial metagenomics

Mitochondria DNA is abundant in animal cells, untargeted NGS of arthropods samples around 1% of mitochondrial reads.



Mitochondrial metagenomics

(or mito-metagenomics or mitogenomics...)

Mitochondria DNA is abundant in animal cells, untargeted NGS of arthropods samples around 1% of mitochondrial reads.

Community ecology

Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification

Xin Zhou^{1,2,3,*}, Yiyuan Li^{1,2,3}, Shanlin Liu^{1,2,3}, Qing Yang¹, Xu Su^{1,2,3}, Lili Zhou^{1,2}, Min Tang^{1,2}, Rabei Fu¹,

Oxford Journals > Science & Mathematics > Nucleic Acids Research > Volume 42, Issue

Multiplex sequencing of pooled mitochondrial genomes—a crucial step toward biodiversity analysis using mito-metagenomics

Min Tang¹, Meihua Tan^{1,2}, Guanliang Meng^{1,3}, Shenzhou Yang¹, Xu Su¹, Shanlin Liu¹, Wenhui Song¹, Yiyuan Li¹, Qiong Wu¹, Aibing Zhang⁴ and Xin Zhou^{1,*}

Biodiversity recovery

Bulk De Novo Mitogenome Assembly from Pooled Total DNA Elucidates the Phylogeny of Weevils (Coleoptera: Curculionoidea)

Conrad P.D.T. Gillett^{1,2}, Alex Crampton-Platt^{1,3}, Martijn J.T.N. Timmermans^{1,4}, Bjarte

Soup to Tree: The Phylogeny of Beetles Inferred by Mitochondrial Metagenomics of a Bornean Rainforest Sample

Alex Crampton-Platt^{1,2}, Martijn J.T.N. Timmermans^{1,3}, Matthew L. Gimmel⁴, Sujatha Narayanan Kuttu^{2,1}, Timothy D. Cockerill^{1,5}, Chey Yun Khen⁶ and Alfried P. Vogler^{1,3,*}

Large-scale phylogenetics

Phylogenetic community ecology of soil biodiversity using mitochondrial metagenomics

Carmelo Andújar^{1,2,*}, Paula Arribas^{1,2}, Filip Ruzicka^{1,3}, Alex Crampton-Platt^{1,3}, Martijn J.T.N. Timmermans^{1,2,†} and Alfried P. Vogler^{1,2}

Issue



Molecular Ecology

Validating the power of mitochondrial metagenomics for community ecology and phylogenetics of complex assemblages

Carola Gómez-Rodríguez^{1,2,*}, Alex Crampton-Platt^{1,3}, Martijn J. T. N. Timmermans^{1,4,5}, Andrés Baselga² and Alfried P. Vogler^{1,4}

Issue



Methods in Ecology and Evolution

Volume 6, Issue 8, pages 883–894, August 2015

Pollinators Monitoring

High-throughput monitoring of wild bee diversity and abundance via mitogenomics

Min Tang^{1,†}, Chloe J. Hardman^{2,†}, Yinqiu Ji^{3,†}, Guanliang Meng¹, Shanlin Liu¹, Meihua Tan^{1,4}, Shenzhou Yang¹, Ellen D. Moss⁵, Jiaxin Wang³, Chenxue Yang³, Catharine Bruce⁶, Tim Nevard^{7,8}, Simon G. Potts², Xin Zhou^{1,*} and Douglas W. Yu^{3,6,*}

Issue

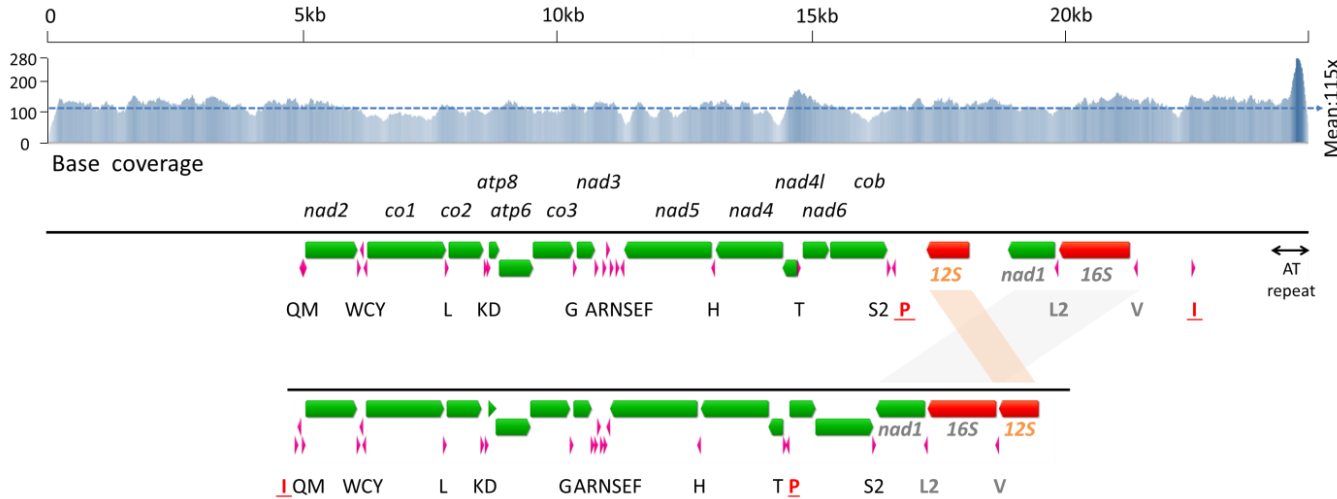


Methods in Ecology and Evolution

Volume 6, Issue 9 1034–1043, Septe

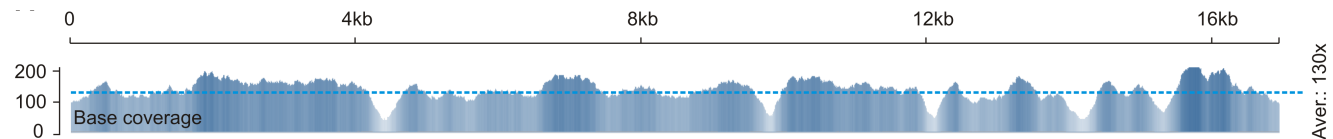
(parenthesis) Mitochondrial genome evolution

Detection of complex mitochondrial rearrangements that explain some Barcoding failures



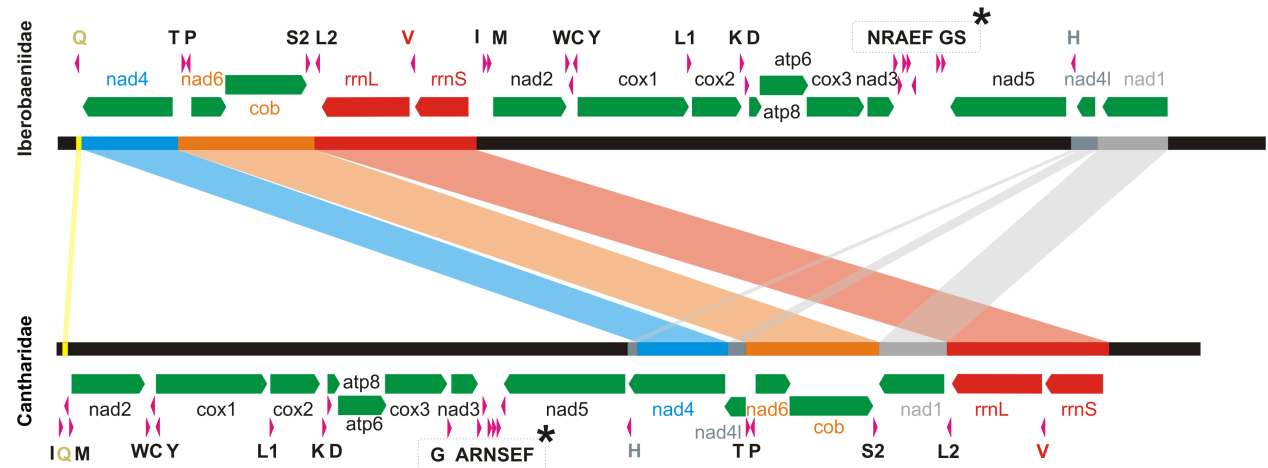
Hydropsyche pellucidula
(Hydropsychidae)
first gene rearrangement in the
insect order Trichoptera

Linard B, et al.
Mitochondrial DNA, 2015











Iberobaenia
(Coleoptera: iberobaeniidae)
First rearrangement of
protein coding genes in the
beetles

Andújar C et al.
(in review)



Skimming arthropod communities

Skimming an animal community will use the same sampling principles (high-copy motifs first) but has also specific characteristics ... **it's a META Genomic Skimming (MGS)**

	Genome skimming on plant pools	Meta-mitogenomics on insect pools
PCR-free sequencing (organelles & genome)		
Shallow sequencing		
Strategy	(Generally) Multiplexed sequencing	1 extraction for the whole pool, Anonymous reads
# of morphospecies	1	Many
Phyletic diversity	Low	High
Genome complexity (per specimen)	High	Low
Chloroplast		
Gut content		

Problematic

Our approach is similar to previous Genome Skimming works,
but with specific characteristics: “Metagenome skimming” (MGS) ...

Which genomic elements are extensively sampled from an arthropod metagenome ?

Can we recover information from gut contents ?

Can MGS teach us something about arthropod genomes and evolution ?

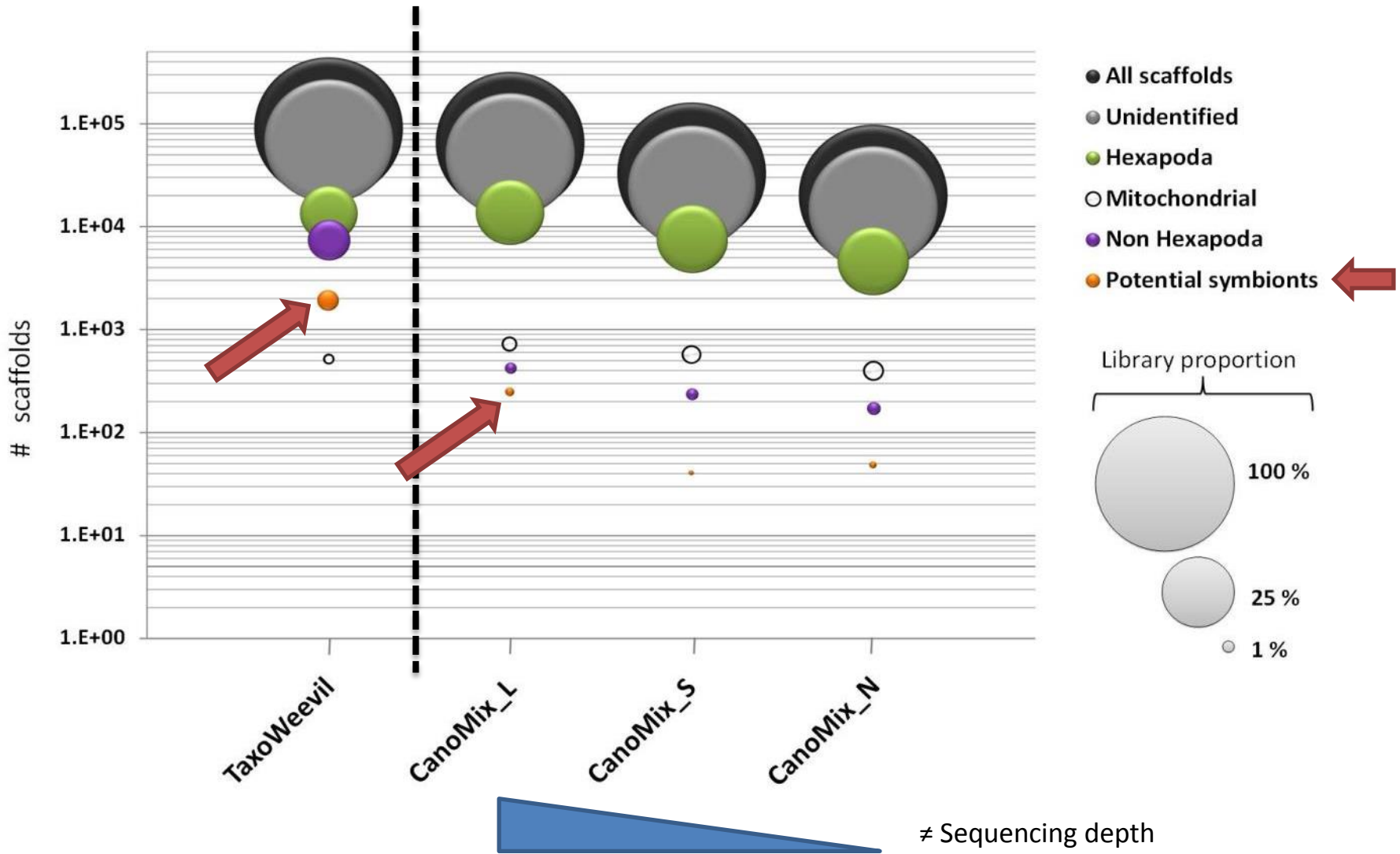


Samples: Field capture or taxon assemblage

Sample	Content			Read pair
	Specimens	Morpho-species	Super-families	
TaxoWeevils	173	173	1	17389929
CanoMix	480	212	14	23922520

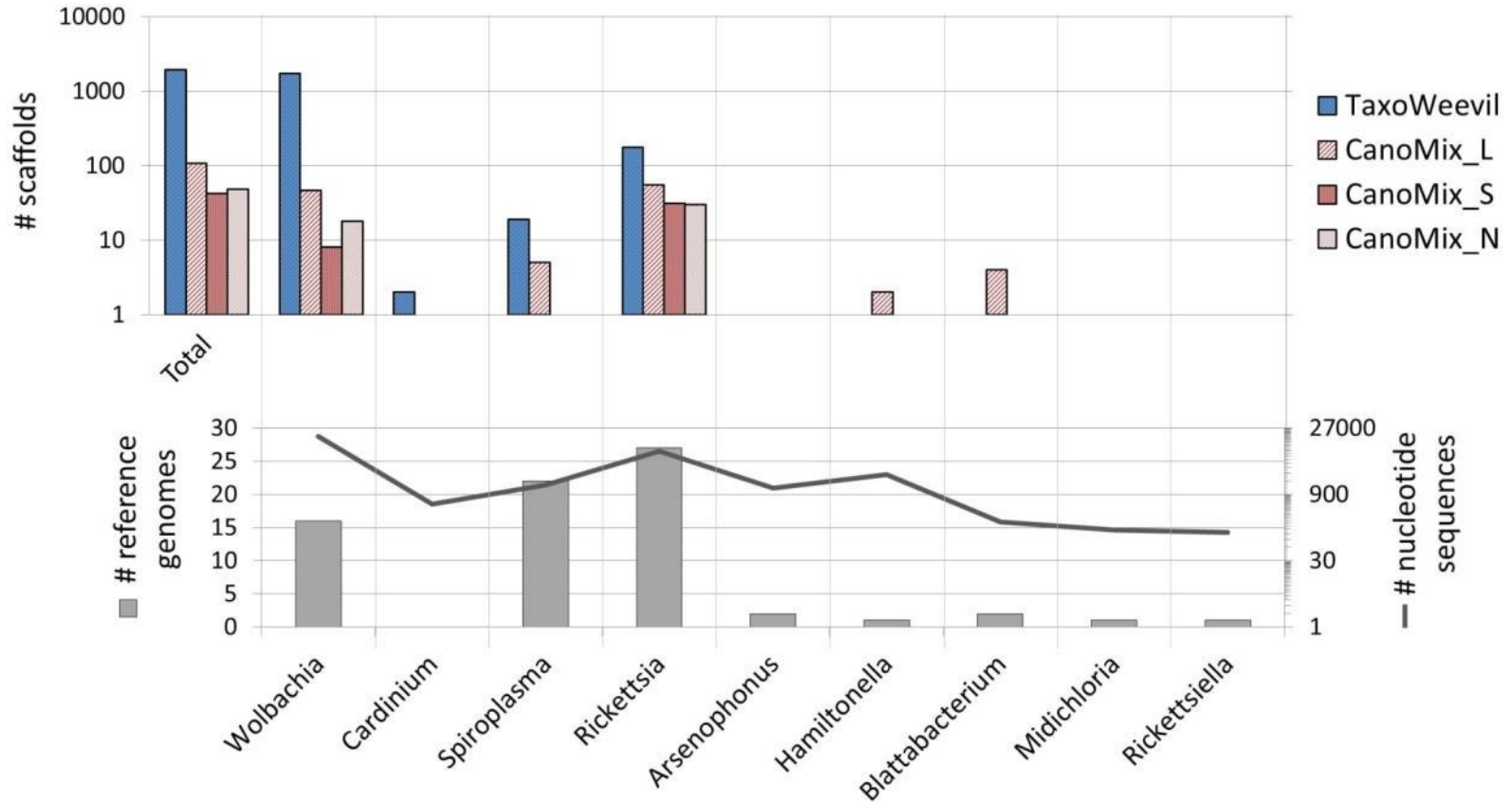
DNA Scaffolds identification

- Annotation by homology to 3 complete NCBI databases (nt, est, genomes)
- Categorized by their best blast hits



Metagenome skimming and associated fauna

- Large sampling of associated bacterial symbionts :



- (Canopy sample only) cherry on the cake ... Plastids and rRNAs scaffolds

Diet remnants ?

- * Dozens of *Coccoloba* family chloroplasts in Canopy sample

Parasites ?

- * 7kb nematode scaffold
- * rRNA 99.5% similar to *Glarea lozoyensis* (fungi, insect pathogen)

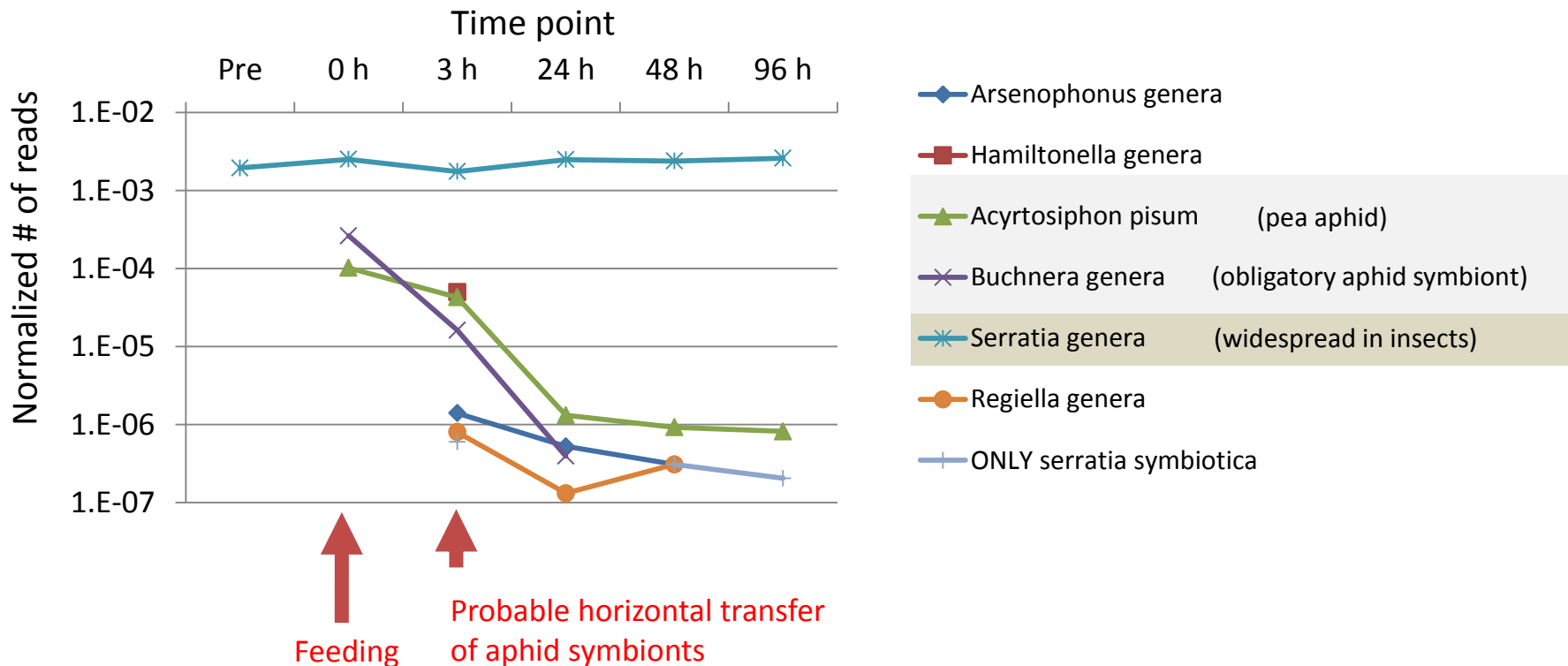
Detecting bacterial symbionts via total DNA extraction

→ [Harmonia axyridis](#) (coleoptera:coccinellidae)

- Lady bird fed with 1 aphid at 0 h
- DNA decay monitored in the gut by shallow sequencing at different time points



Collaboration with
Debora Pires-Paula,
*Embrapa Genetic Resources
and Biotechnology, Brazilia*

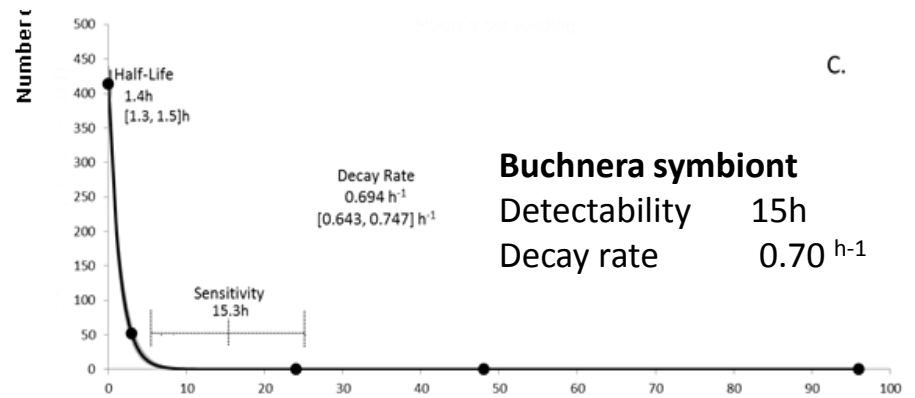
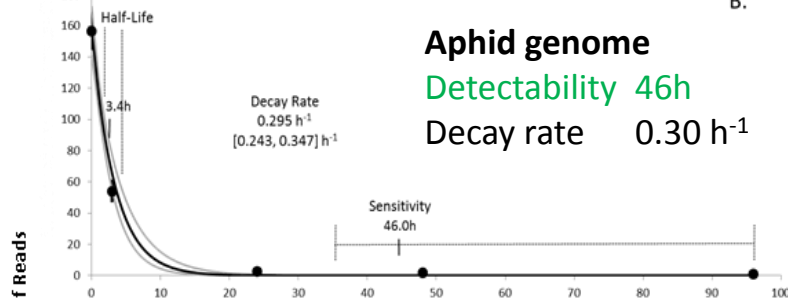
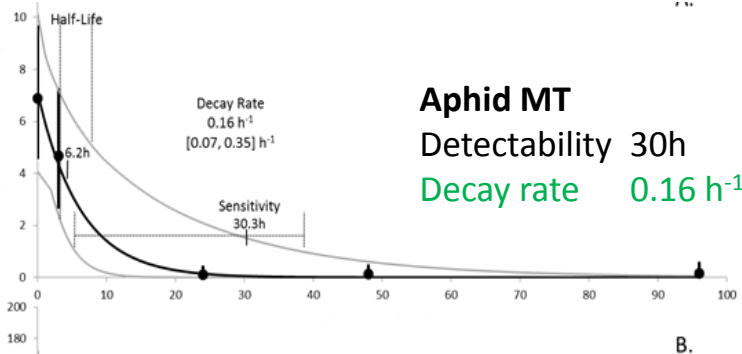


Bayesian models of food decay

	Elapsed time after feeding					
	Pre	0 h	3 h	24 h	48 h	96 h
A. pisum mtDNA	0	27	15	0	0	0
A. pisum nuclear DNA	0	624	214	10	6	4
B. aphidicola	0	1,651	171	2	0	0

Detection and decay rates of prey and prey symbionts in the gut of a predator through metagenomics

Débora P. Paula^{1,2,*}, Benjamin Linard²,
David A. Andow³, Edison R. Sujji¹,
Carmen S. S. Pires¹ and Alfried P. Vogler^{2,4}



Gut content used for intraguild predation analysis

Debora's second project Trophic interaction between carnivorous Ladybirds
From Brazilian agrocultures



Coccinellini

*Cycloneda
sanguinea*

Coccinellini

*Harmonia
axyridis*



Doru luteipes

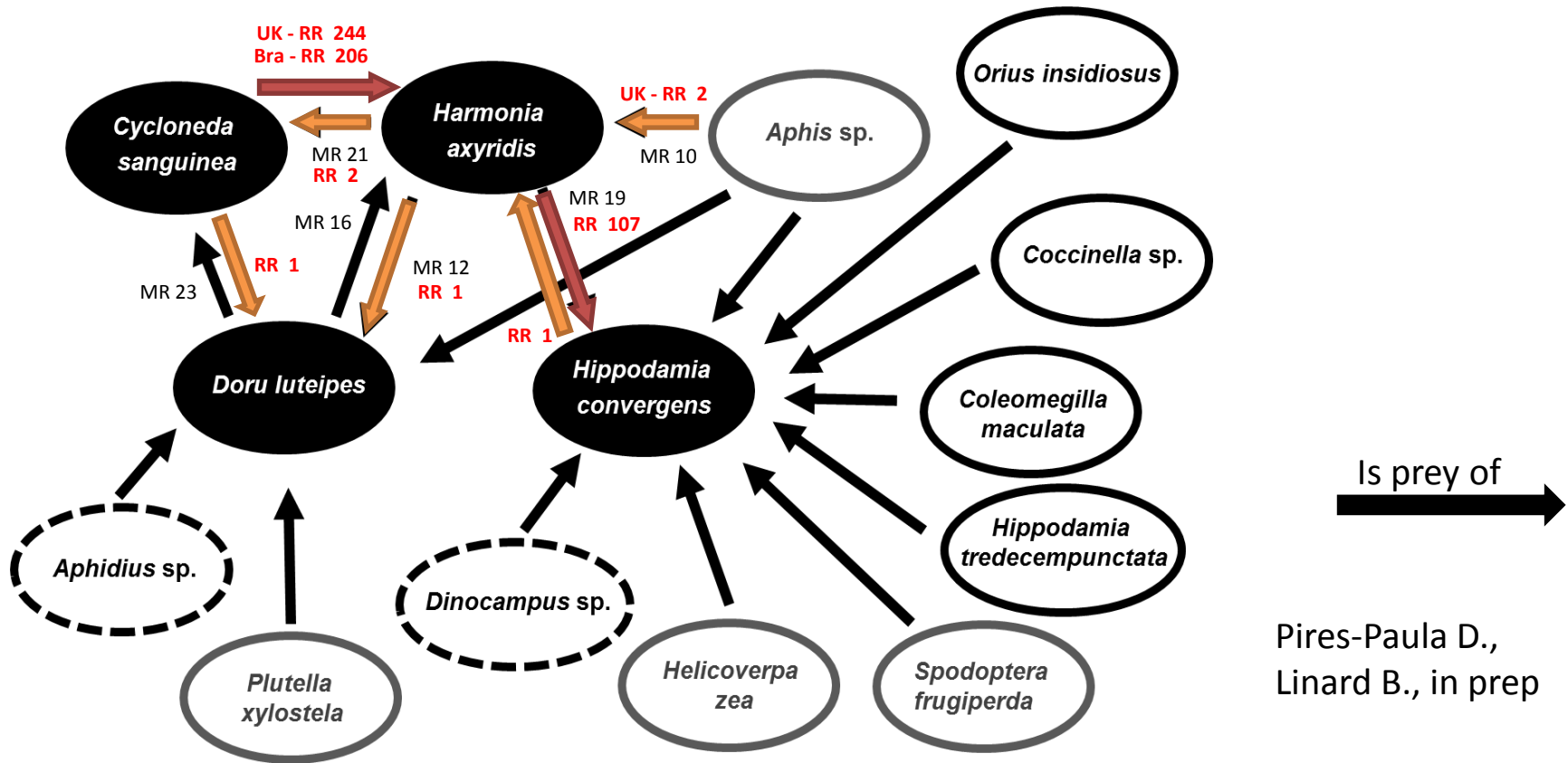
Earwig
(Dermaptera)

*Hippodamia
convergens*

Coccinellini



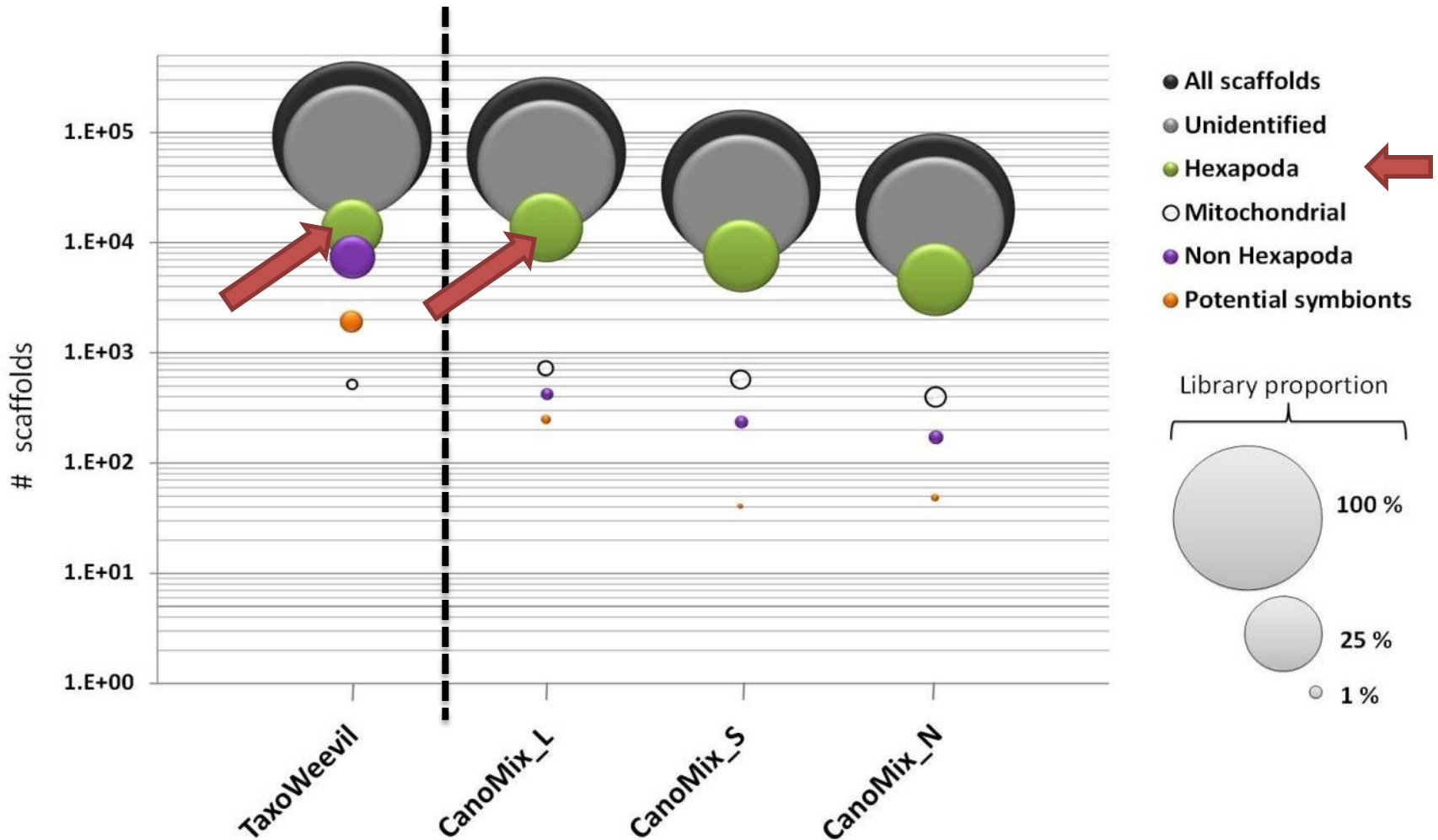
Building an exploratory food web



Pires-Paula D.,
Linard B., in prep

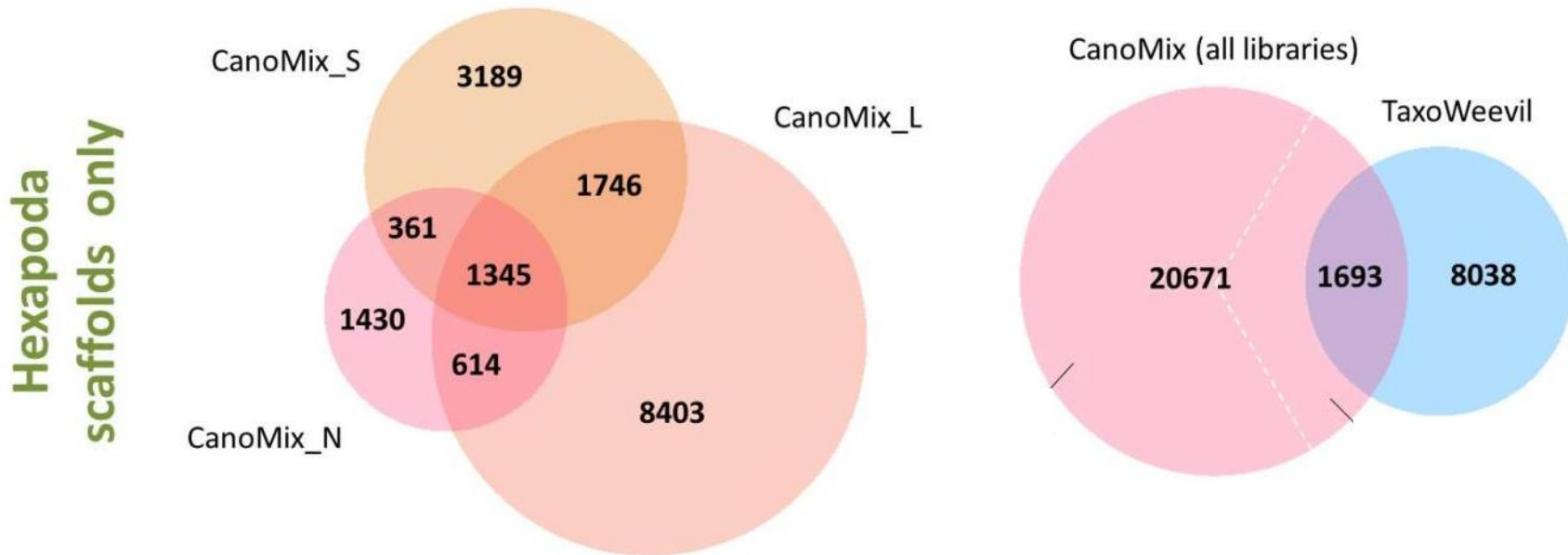
DNA Scaffolds identification

- Annotation by homology to 3 complete NCBI databases (nt, est, genomes)
- Categorized by their best blast hits



Nature of Hexapoda scaffolds

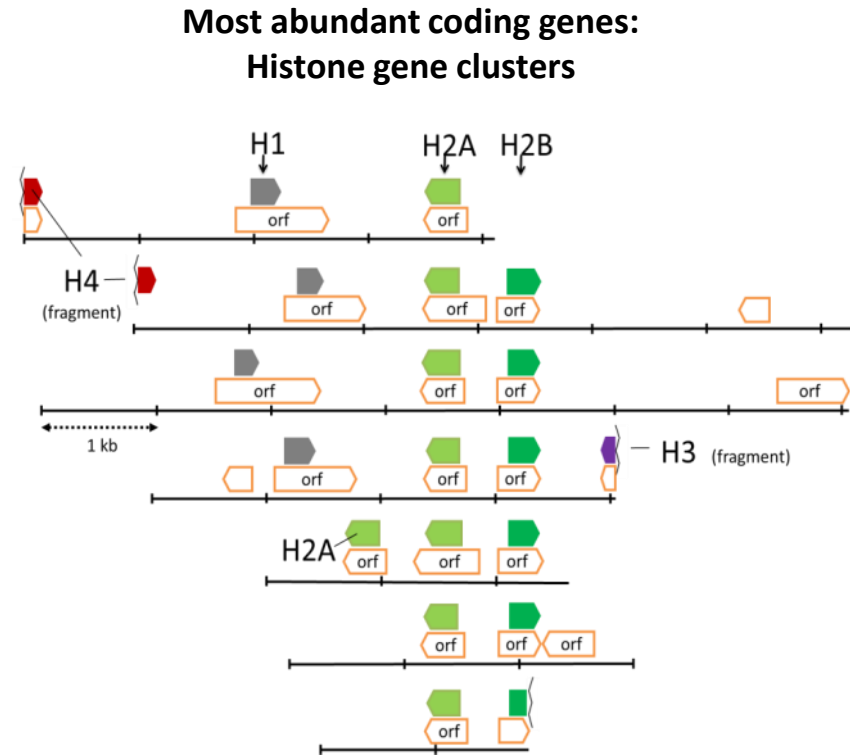
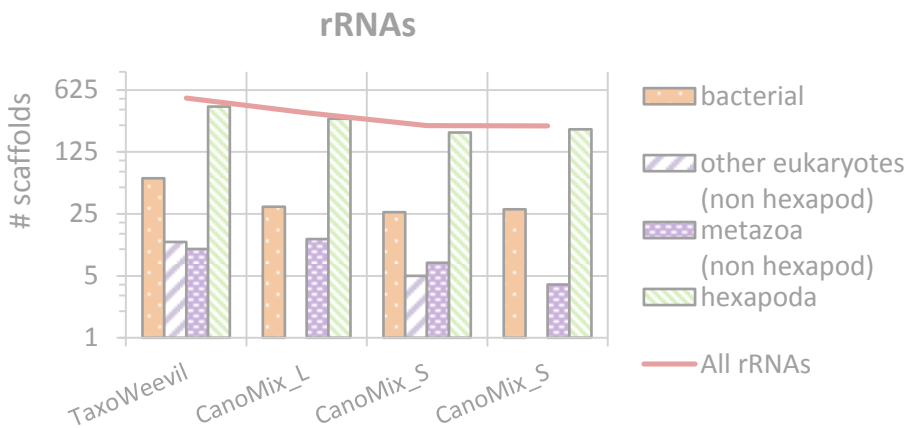
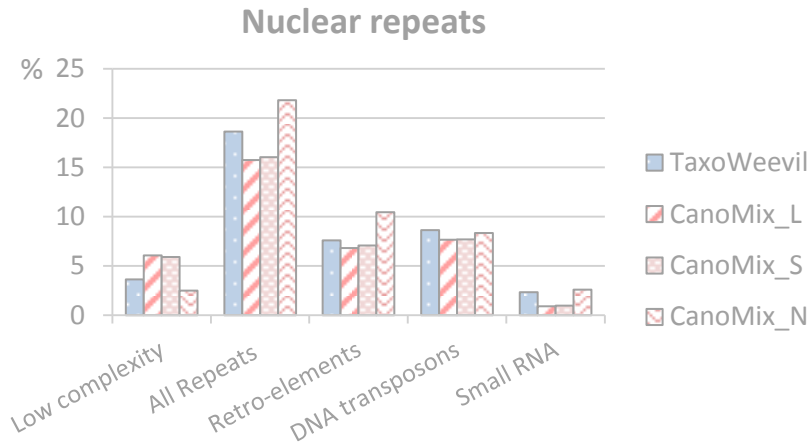
- Library intersections:



→ A core set of sequences is systematically recovered

Nature of Hexapoda scaffolds

- Library intersections: → **A core set of sequences is systematically recovered**
- Nature of the sequences: during MGS, we expect to sample multicopy elements.



→ **Potential phylogenetic marker**
(Talbert et al. 2012)

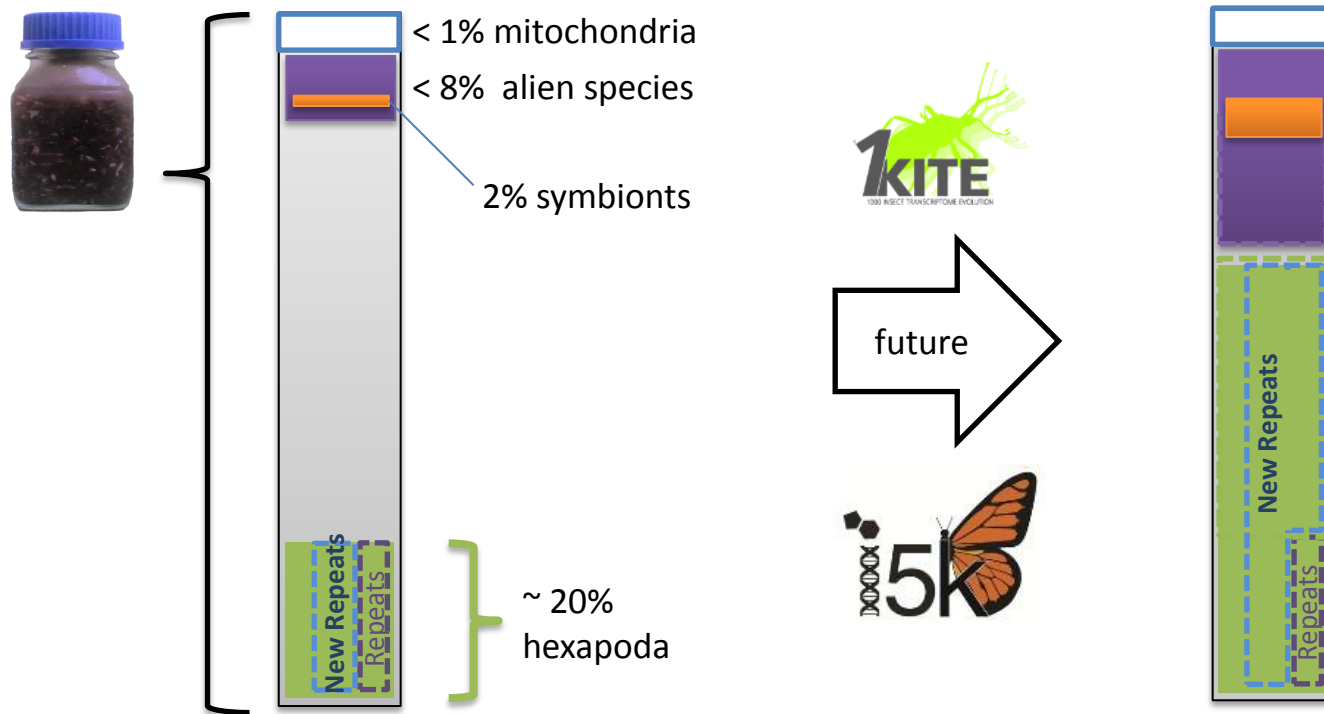
→ **Gene sequence + cluster rearrangements**

Metagenome skimming of arthropod specimen pools

What we identify in one pool:

Metagenome Skimming of Insect Specimen Pools: Potential for Comparative Genomics

Benjamin Linard¹, Alex Crampton-Platt^{1,2}, Conrad P.D.T. Gillett¹,
Martijn J.T.N. Timmermans^{1,3} and Alfried P. Vogler^{1,3,*}



- Up to 42% of insect genomes are repetitive (Wang et al. 2008)
- Less than 0.15% of repeats found to be homologous between the two coleopteran genomes sequenced to date (Keeling et al. 2013)

→ **clade-specific nuclear repeats are the most sampled sequences**

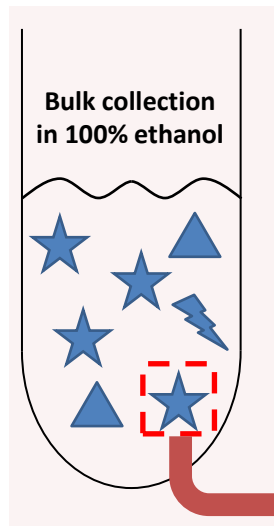
Metagenome skimming of **preserving ethanol**

MGS is semi-destructive, what if new undescribed species were collected ?

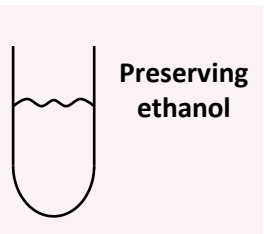
Wait... all our communities are collected in 100% ethanol !

Hajibabei et al., 2012 -> high species recovery via metabarcoding of preserving ethanol

A. Arthropod sample



C. Ethanol metagenome skimming (ethaMGS)



Total DNA precipitate

NGS sequencing

Taxonomic analysis of all reads

MITOCHONDRIAL READS

CONCOMITANT DNA

Associated fauna & ecological traits

Species & biomass recovery

B. Mitochondrial Metagenomics (MMG)



Independent extraction of total DNA

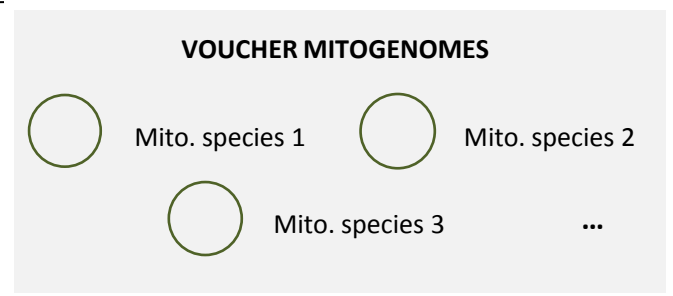
Equimolar pooling

NGS sequencing

COX1 barcoding (Independent Sanger sequencing)

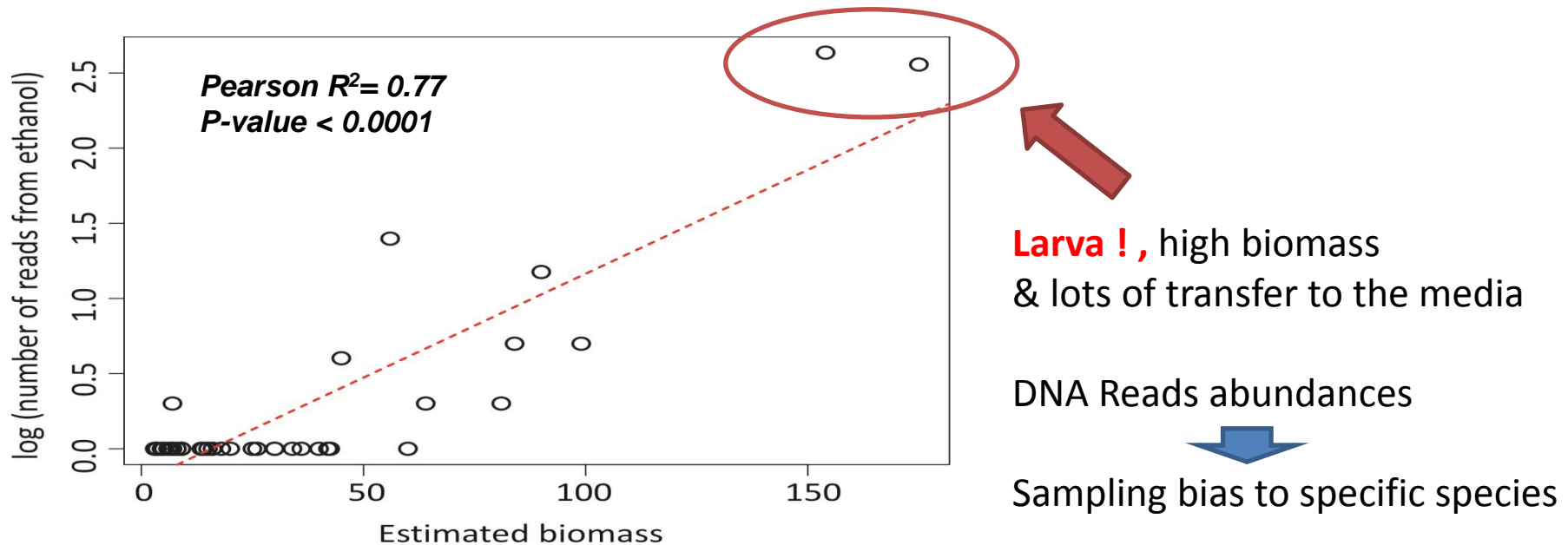
Mitochondrial contigs (reads assembly)

Sequence match



Metagenome skimming of preserving ethanol

	Vouchers (MMG)		Ethanol reads (ethaMGS)			
	Recovery	cox1 Sanger	Mitogenome	matching cox1	matching complete mitogenome	matching prot-coding regions
Aquatic (21 species)	Species	20	21	2	9	7
	Proportion	95%	100%	9.5%	43%	33%
Ground (19 species)	Species	17	18	2	6	6
	Proportion	89%	95%	11%	32%	32%

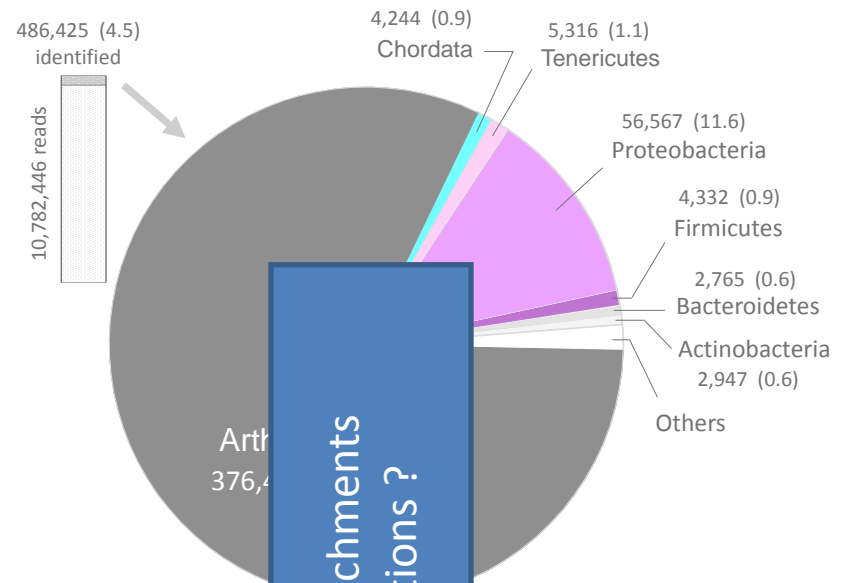
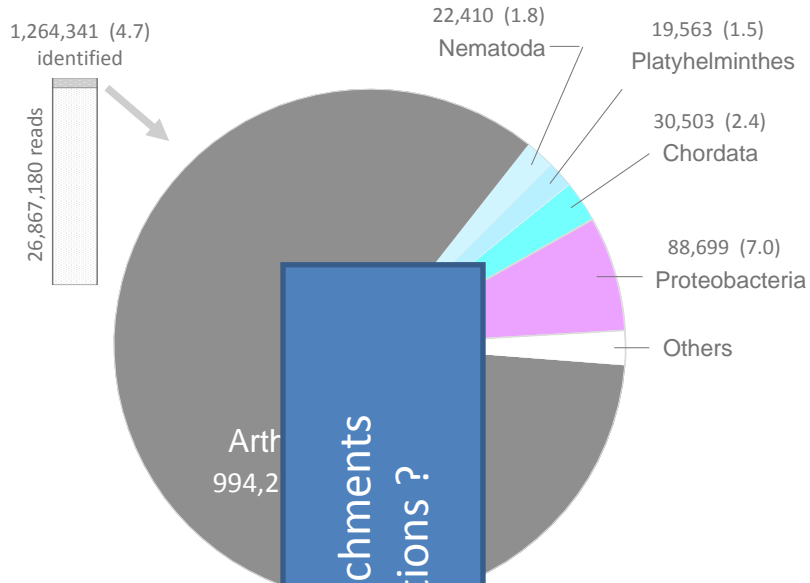




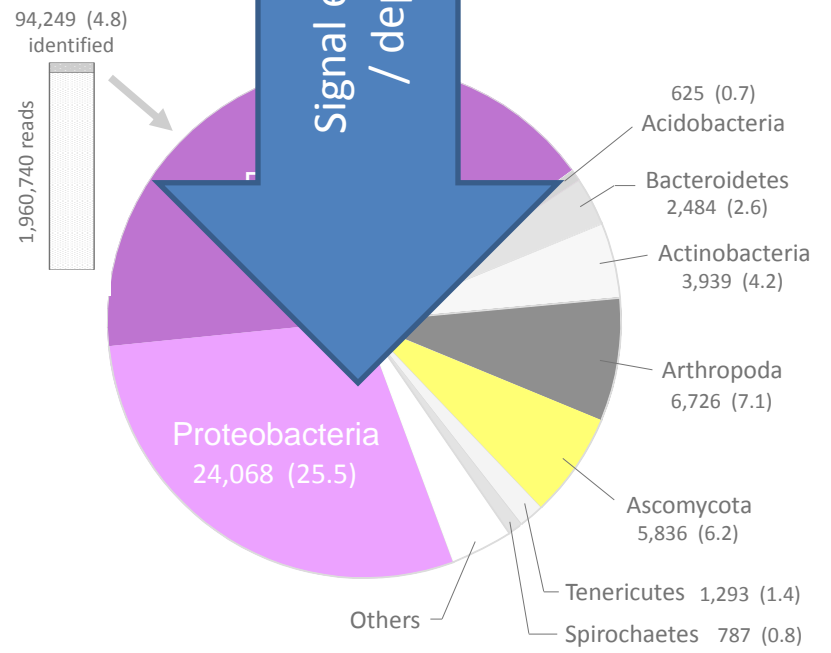
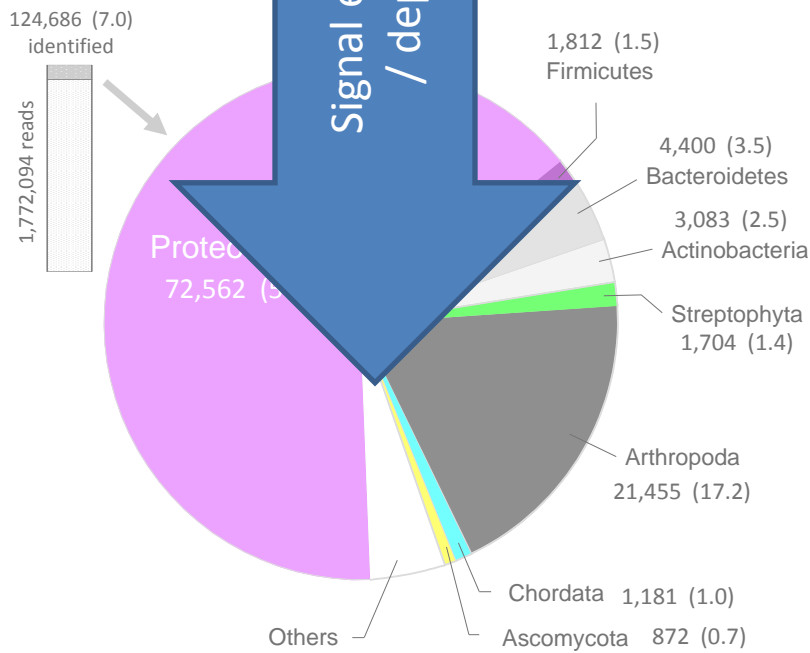
Voucher

Aquatic

Terrestrial



Ethanol



General patterns of concomitant DNAs

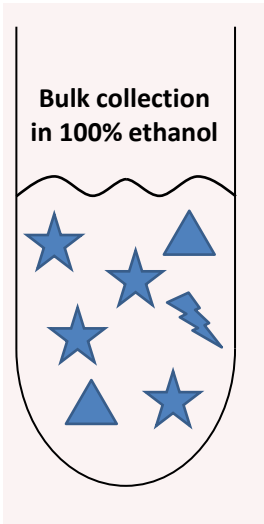
- deep taxonomic analysis of mitochondrial, rRNA, chloroplastic and symbiotic markers.
- Understanding the potential of the ethanol before going back to targeted approaches

Clade	Marker	Aquatic			Terrestrial			Comments
		V	E	$\Delta \text{Fold}_{E/V}$ (log)	V	E	$\Delta \text{Fold}_{E/V}$ (log)	
Arthropoda	Mito			1.9 ▼			2.0 ▼	
	rRNA			2.0 ▼			2.0 ▼	
<i>Annelides</i>	rRNA			4.6 ▲				>99% similar to Enchytraeidae and Naididae, found in benthic and wet soil habitats ^{a,b}
<i>Fungi</i>	Mito						1.8 ▼	In TE, 75% of mito. reads are >99% similar to Metarhizium, an entomopathogen genera ^c
	rRNA			---			2.2 ▲	
<i>Viridiplantae</i>	Plastid			3.7 ▲			3.2 ▲	
	Mito			3.5 ▲			3.0 ▲	
	rRNA			4.6 ▲			---	
<i>Stramenophiles</i>	Plastid			3.3 ▲			3.2 ▲	
	Mito			2.3 ▲			2.9 ▲	
<i>Blastocystis</i>	Mito			4.1 ▲			3.2 ▲	Insect gastrointestinal tracts habitat ^d
<i>Bacteria</i>	<i>Acinetobacter</i>						1.8 ▼	Soil mineralization and found in beetle guts ^{e,f}
	<i>Hydrogenophaga</i>			2.7 ▲				Oxygenates-rich water habitats ^g
	<i>Variovorax</i>			2.8 ▲				Soil and water habitats ^{h,i}

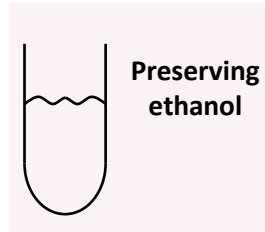
Ethanol MGS : lessons learned

(manuscript in preparation)

Arthropod sample



Ethanol metagenome skimming (ethaMGS)



**Think twice before throwing the ethanol !
It contains ecological traits.**

1. Understanding which signal could hold your community of interest → metagenomics, deep taxonomic analysis

Warning: Larva vs adults, biomass influence, pooling design

2. Eventually, use the ethanol to add value to your study by using targeted approaches (ethanol metabarcoding)

3. If you are interested in symbionts, only open associations are likely to be transferred to the media (vomit effect)

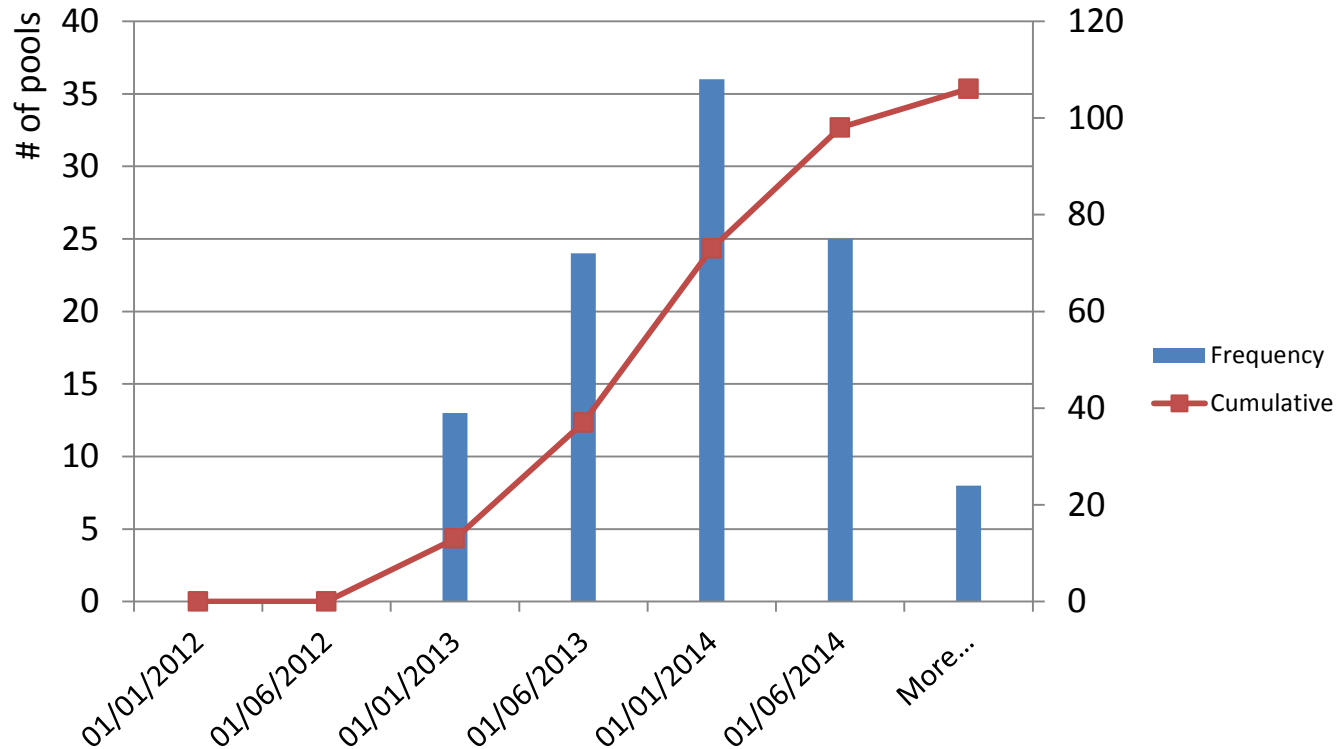


Hey ! You trashed our lunch and symbionts !
Stupid postdoc !

Perspectives: integrative analysis of insect pools

1. A collection of pools ? Some challenges...

→ Pools from different projects and build for different purposes



MODELING : predicting sequencing outcomes

My question: I do a low coverage metagenome of a specimen pool.
(Mito-metagenomics, metagenome skimming...)

What read outcome should I expect

1. for different targets (rRNA, mito, symbionts, histones, repeats...)
2. using a given sequencing depth (# bp sequenced)
3. on a pool of given taxonomic diversity (# specimens/diversity)

Inspired from the future
Directions described in
the review

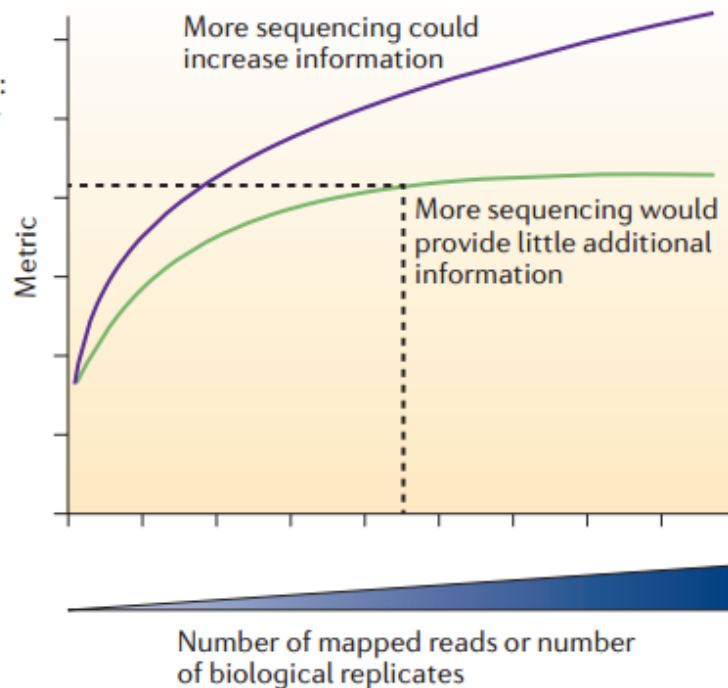
Sequencing depth and coverage:
key considerations in genomic
analyses

David Sims, 2014, Nature reviews

Box 3 | Staged sequencing for predicting sequencing requirements

Possible metrics:

- General transcriptome coverage: percentage of genes covered over 90% at a given expression level
- Differential expression: number of differentially expressed genes
- Alternative isoform detection: percentage of split reads (that is, junction that spans reads)
- ChIP-seq peak detection: number of enriched loci



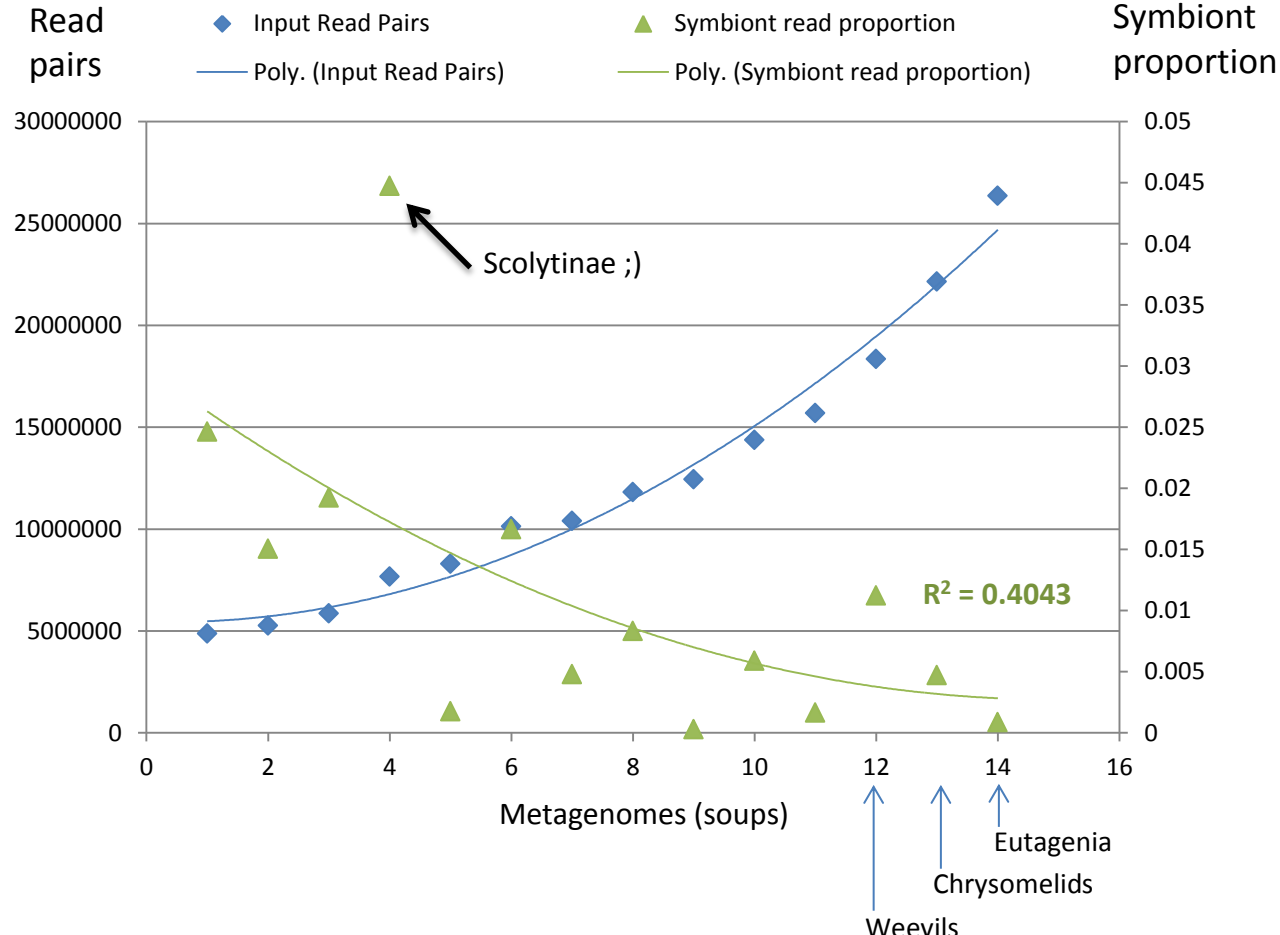
Read level MODELING : some symbiont results

My data : Reads of all libraries mentioned previously.
High thresholds for IDENTIFICATION (>99% identity on 90% of the read)

Very basic example:
(not well normalized)

Will be completed
with the taxo diversity
of the samples...

Lower sequencing depth
= more symbionts
sampled ???



Conclusions

For questions related to species-rich and relatively unknown clades, metagenomic approaches based on organelles show lots of potential.

Mitochondrial mitogenomics can be seen as a « superbarcoding » needing very few wet lab work to build a large genomic reference database.

When enrichment technics will be optimized, thousands of mitochondria per run

Genomics and « large-scale » comparative genomics will be the new core of environmental studies, but new challenges and bioinfo developments are now needed, even for a « well-known » marker like the mitochondria.



Hey ! You trashed our lunch and symbionts !
Stupid postdoc !



All NHM Biodiversity Initiative
members

&

Kevin Hoptkins for NGS advices and support
Peter Foster for bioinfo support
Paul Ward for IT infrastructure

&

NHM/Imperial molecular labs
NHM/Oxford sequencing platforms

Alfried Vogler
Alex Crampton-Platt
Martijn Timmermans
Conrad Gillett
Carmelo Andujar
Chris Barton
Belen Arias
Hannah Norman
Chris Phipps

Paula Arribas Blázquez
Kirsten Miller
Debora Pires Paula
Rosli Kasah
Emeline Favreau
Carola Gomez
Andres Baselga
CT Tang



Thank you for your attention.