

The number of physiologically important microRNA targets has been vastly overestimated

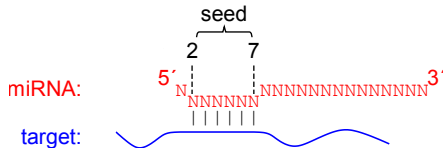
Hervé Seitz

October 18, 2024

Slideshow accessible for download at:

<http://www.igh.cnrs.fr/equip/Seitz/EMB024.pdf>

miRNA target identification



Overestimated
number of miRNA
targets

miRNA target
identification

False positives in
target prediction

Binding site
evolution after
miRNA loss

Strategy

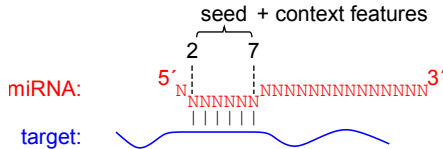
Observed losses

Evolution of seed match
count after loss

Conclusion

Supplementary
data

miRNA target identification



Overestimated
number of miRNA
targets

miRNA target
identification

False positives in
target prediction

Binding site
evolution after
miRNA loss

Strategy

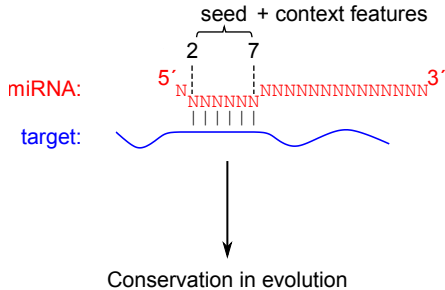
Observed losses

Evolution of seed match
count after loss

Conclusion

Supplementary
data

miRNA target identification



Overestimated
number of miRNA
targets

miRNA target
identification

False positives in
target prediction

Binding site
evolution after
miRNA loss

Strategy

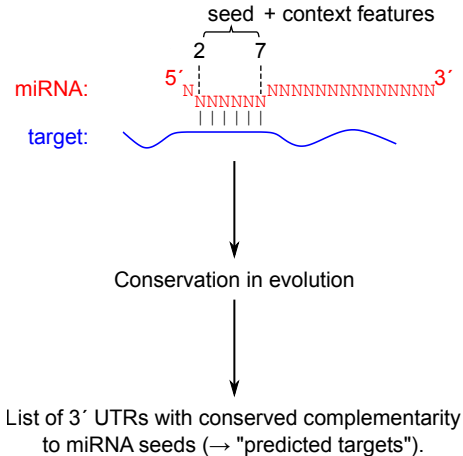
Observed losses

Evolution of seed match
count after loss

Conclusion

Supplementary
data

miRNA target identification



Overestimated
number of miRNA
targets

miRNA target
identification

False positives in
target prediction

Binding site
evolution after
miRNA loss

Strategy

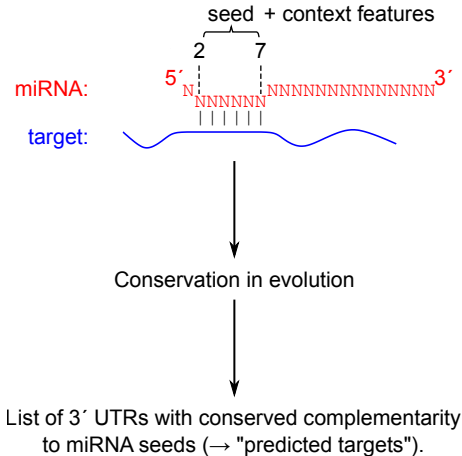
Observed losses

Evolution of seed match
count after loss

Conclusion

Supplementary
data

miRNA target identification



Even high-throughput identification needs to assess conservation.

Overestimated
number of miRNA
targets

miRNA target
identification

False positives in
target prediction

Binding site
evolution after
miRNA loss

Strategy

Observed losses

Evolution of seed match
count after loss

Conclusion

Supplementary
data

False positives in target prediction

Some 3' UTR seed matches are conserved for miRNA-independent reasons.

Overestimated
number of miRNA
targets

miRNA target
identification

False positives in
target prediction

Binding site
evolution after
miRNA loss

Strategy

Observed losses

Evolution of seed match
count after loss

Conclusion

Supplementary
data

False positives in target prediction

Overestimated
number of miRNA
targets

miRNA target
identification

False positives in
target prediction

Binding site
evolution after
miRNA loss

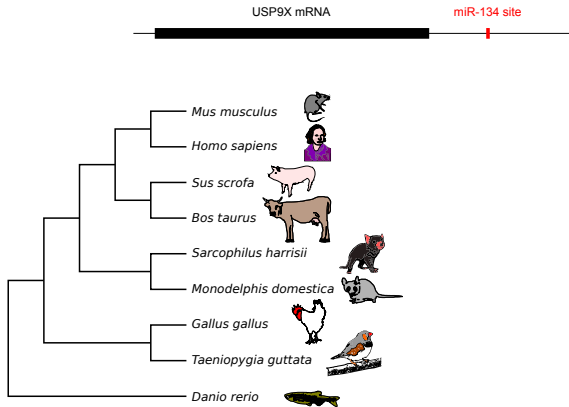
Strategy

Observed losses

Evolution of seed match
count after loss

Conclusion

Supplementary
data



(Pinzón *et al.*, 2017)

False positives in target prediction

Overestimated number of miRNA targets

miRNA target identification

False positives in target prediction

Binding site evolution after miRNA loss

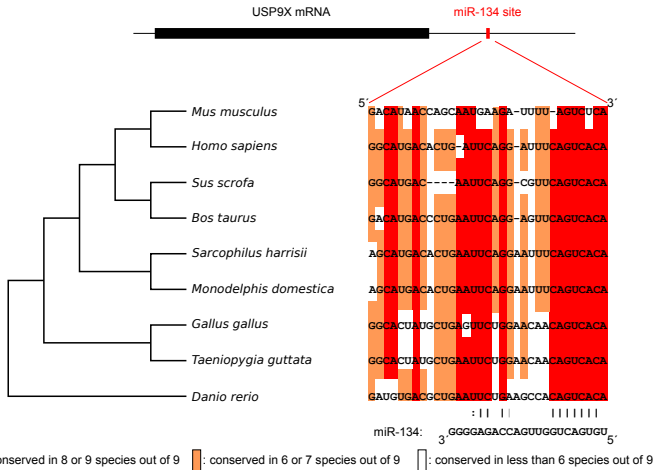
Strategy

Observed losses

Evolution of seed match count after loss

Conclusion

Supplementary data



(Pinzón *et al.*, 2017)

False positives in target prediction

Overestimated number of miRNA targets

miRNA target identification

False positives in target prediction

Binding site evolution after miRNA loss

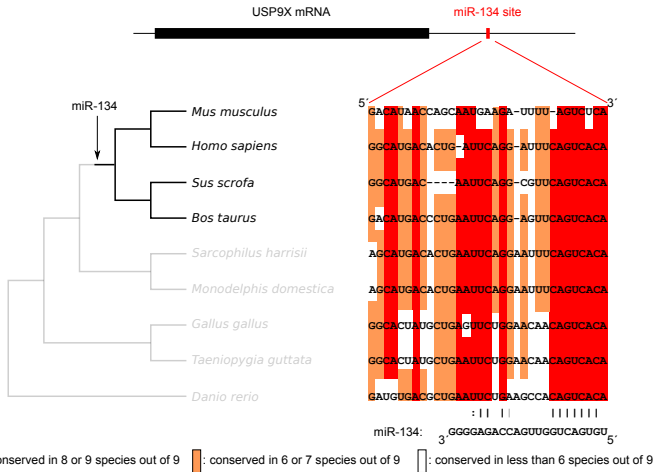
Strategy

Observed losses

Evolution of seed match count after loss

Conclusion

Supplementary data



(Pinzón *et al.*, 2017)

False positives in target prediction

Measuring miRNA-independent conservation with control k -mers (with same G/C content, dinucleotide composition, ...; [Friedman *et al.*, 2009](#)).

Overestimated
number of miRNA
targets

miRNA target
identification

False positives in
target prediction

Binding site
evolution after
miRNA loss

Strategy

Observed losses

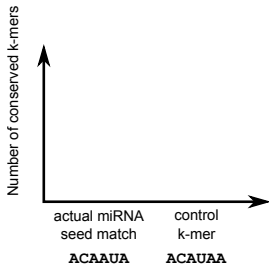
Evolution of seed match
count after loss

Conclusion

Supplementary
data

False positives in target prediction

Measuring miRNA-independent conservation with control k -mers (with same G/C content, dinucleotide composition, ...; [Friedman *et al.*, 2009](#)).



Overestimated number of miRNA targets

miRNA target identification

False positives in target prediction

Binding site evolution after miRNA loss

Strategy

Observed losses

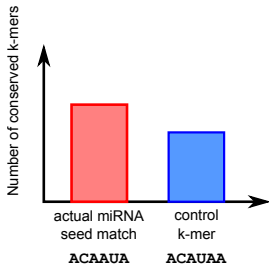
Evolution of seed match count after loss

Conclusion

Supplementary data

False positives in target prediction

Measuring miRNA-independent conservation with control k -mers (with same G/C content, dinucleotide composition, ...; [Friedman *et al.*, 2009](#)).



Overestimated
number of miRNA
targets

miRNA target
identification

False positives in
target prediction

Binding site
evolution after
miRNA loss

Strategy

Observed losses

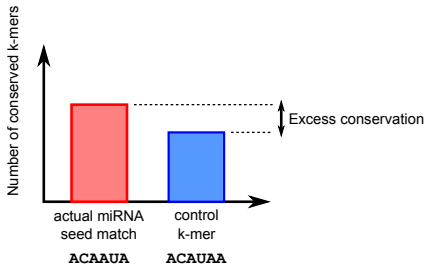
Evolution of seed match
count after loss

Conclusion

Supplementary
data

False positives in target prediction

Measuring miRNA-independent conservation with control k -mers (with same G/C content, dinucleotide composition, ...; [Friedman *et al.*, 2009](#)).



Overestimated
number of miRNA
targets

miRNA target
identification

False positives in
target prediction

Binding site
evolution after
miRNA loss

Strategy

Observed losses

Evolution of seed match
count after loss

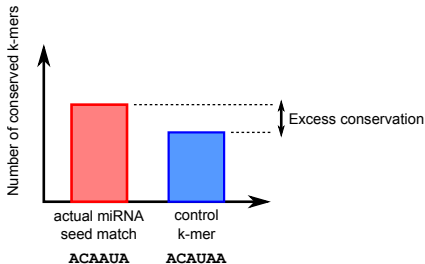
Conclusion

Supplementary
data

False positives in target prediction

Measuring miRNA-independent conservation with control k -mers (with same G/C content, dinucleotide composition, ...; [Friedman *et al.*, 2009](#)).

→ not submitted to the same selective pressure than actual miRNA seed matches (e.g., protein binding sites).



Overestimated number of miRNA targets

miRNA target identification

False positives in target prediction

Binding site evolution after miRNA loss

Strategy

Observed losses

Evolution of seed match count after loss

Conclusion

Supplementary data

Binding site evolution after miRNA loss

Overestimated
number of miRNA
targets

miRNA target
identification

False positives in
target prediction

Binding site
evolution after
miRNA loss

Strategy

Observed losses

Evolution of seed match
count after loss

Conclusion

Supplementary
data

Binding site evolution after miRNA loss

After loss of a *complete* miRNA family (*), how long do its seed matches persist ?

(*): family: all the miRNAs sharing the same seed sequence

Overestimated
number of miRNA
targets

miRNA target
identification

False positives in
target prediction

Binding site
evolution after
miRNA loss

Strategy

Observed losses

Evolution of seed match
count after loss

Conclusion

Supplementary
data

Binding site evolution after miRNA loss

After loss of a *complete* miRNA family (*), how long do its seed matches persist ?

(*): family: all the miRNAs sharing the same seed sequence

We need to be lucky: loss of complete miRNA families!

Overestimated
number of miRNA
targets

miRNA target
identification

False positives in
target prediction

Binding site
evolution after
miRNA loss

Strategy

Observed losses

Evolution of seed match
count after loss

Conclusion

Supplementary
data

Strategy

Overestimated
number of miRNA
targets

miRNA target
identification

False positives in
target prediction

Binding site
evolution after
miRNA loss

Strategy

Observed losses

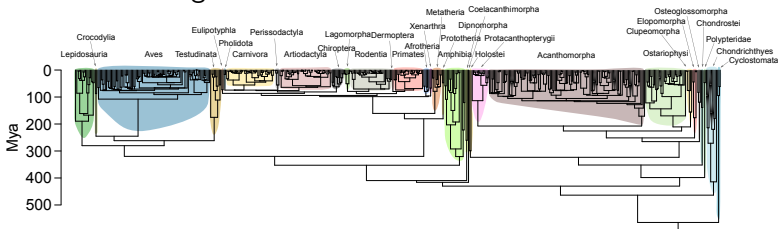
Evolution of seed match
count after loss

Conclusion

Supplementary
data

Strategy

522 vertebrate genomes:



Overestimated
number of miRNA
targets

miRNA target
identification

False positives in
target prediction

Binding site
evolution after
miRNA loss

Strategy

Observed losses

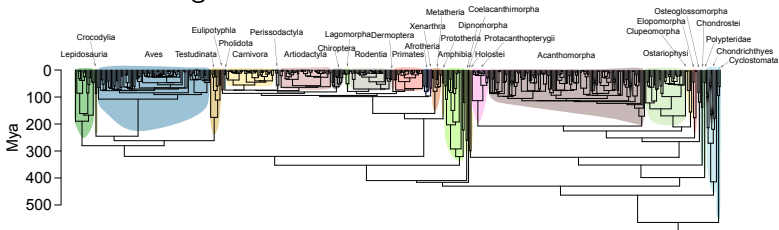
Evolution of seed match
count after loss

Conclusion

Supplementary
data

Strategy

522 vertebrate genomes:



848 best annotated vertebrate miRNA families.

Overestimated
number of miRNA
targets

miRNA target
identification

False positives in
target prediction

Binding site
evolution after
miRNA loss

Strategy

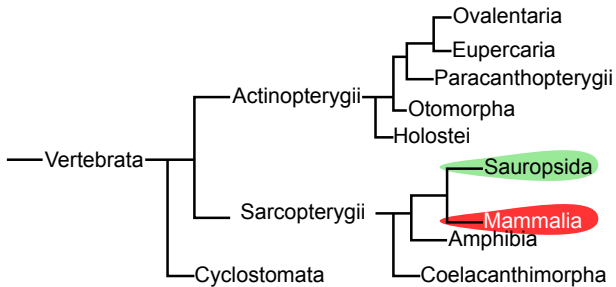
Observed losses

Evolution of seed match
count after loss

Conclusion

Supplementary
data

Strategy



Overestimated
number of miRNA
targets

miRNA target
identification

False positives in
target prediction

Binding site
evolution after
miRNA loss

Strategy

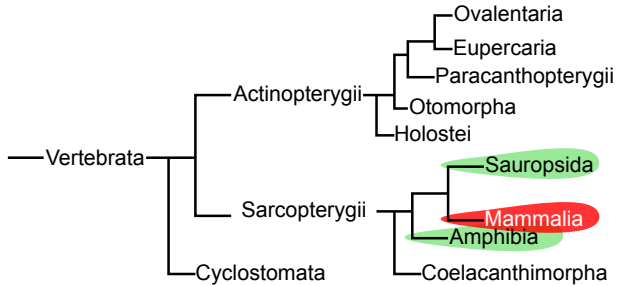
Observed losses

Evolution of seed match
count after loss

Conclusion

Supplementary
data

Strategy



Overestimated
number of miRNA
targets

miRNA target
identification

False positives in
target prediction

Binding site
evolution after
miRNA loss

Strategy

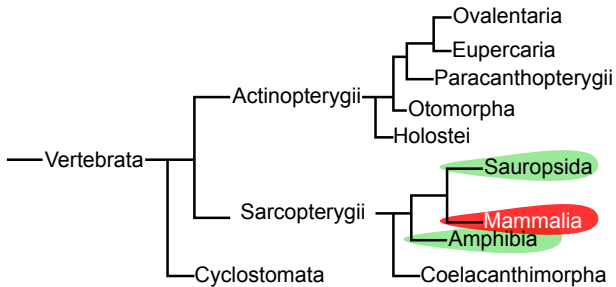
Observed losses

Evolution of seed match
count after loss

Conclusion

Supplementary
data

Strategy



More convincing if many species in clade of loss.

Overestimated
number of miRNA
targets

miRNA target
identification

False positives in
target prediction

Binding site
evolution after
miRNA loss

Strategy

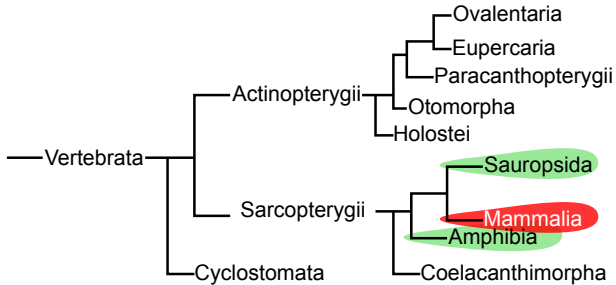
Observed losses

Evolution of seed match
count after loss

Conclusion

Supplementary
data

Strategy



More convincing if many species in clade of loss.

≥ 5 species in clade of loss; presence in sister clade + 3 closest cousin clades.

Overestimated
number of miRNA
targets

miRNA target
identification

False positives in
target prediction

Binding site
evolution after
miRNA loss

Strategy

Observed losses

Evolution of seed match
count after loss

Conclusion

Supplementary
data

Observed losses

Overestimated
number of miRNA
targets

Family	Clade of loss
miR-509	Cervidae+Bovidae
miR-541	Feliformia
miR-224	Mustelidae
miR-338	Lagomorpha
miR-338	Syngnathidae
miR-460	Serpentes
miR-660	Murinae
miR-880	Deomyiinae+Gerbilinae
miR-452	Mustelidae
miR-456	Balistidae+Tetraodontidae
miR-296-3p	Lagomorpha
miR-2188	Eutheria
miR-2114	Phocidae
miR-656	Myomorpha
miR-654	Delphinoida
miR-460b	Mammalia
miR-460b	Serpentes
miR-676	Cervidae+Bovidae
miR-676	Feloidea
miR-187	Serpentes
miR-769	Yangochiroptera
miR-769	Whippomorpha
miR-1185	Yangochiroptera

Family	Clade of loss
miR-451a	Paralichthyidae+Pleuronectidae
miR-509	Yangochiroptera
miR-31	Protacanthopterygii
miR-31	Lophiiformes+Tetraodontiformes
miR-885	Myomorpha
miR-296-5p	Lagomorpha
miR-296-5p	Canidae
miR-412	Lagomorpha
miR-889	Muridae+Cricetidae
miR-889	Bovinae
miR-458	Theria
miR-374	Yangochiroptera
miR-513b	Bos
miR-653	Yangochiroptera
miR-1662	Endoglyptodonta
miR-297p	Pinnipedia+Musteloidea
miR-2131	Psittaciformes
miR-122b	Paralichthyidae+Pleuronectidae
miR-325-3p	Bovidae
miR-599	Batrachia
miR-599	Corvides+Passerides
miR-599	Galloanserae

miRNA target
identification

False positives in
target prediction

Binding site
evolution after
miRNA loss

Strategy

Observed losses

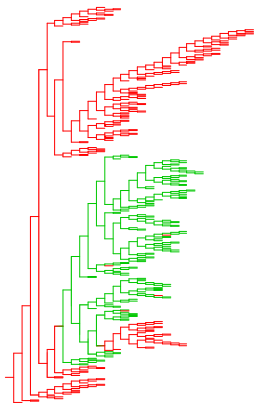
Evolution of seed match
count after loss

Conclusion

Supplementary
data

Observed losses

miR-656 family:



Overestimated
number of miRNA
targets

miRNA target
identification

False positives in
target prediction

Binding site
evolution after
miRNA loss

Strategy

Observed losses

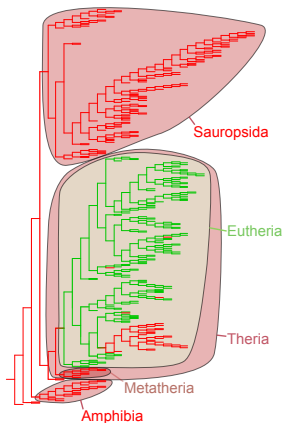
Evolution of seed match
count after loss

Conclusion

Supplementary
data

Observed losses

miR-656 family:



Overestimated
number of miRNA
targets

miRNA target
identification

False positives in
target prediction

Binding site
evolution after
miRNA loss

Strategy

Observed losses

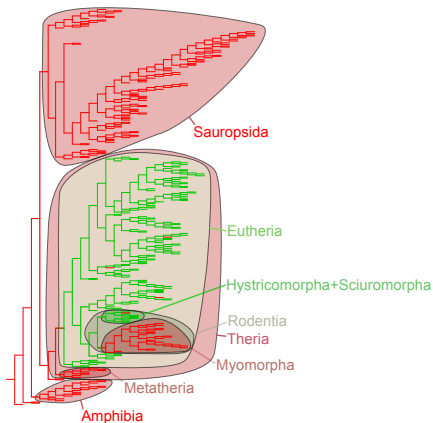
Evolution of seed match
count after loss

Conclusion

Supplementary
data

Observed losses

miR-656 family:



Overestimated
number of miRNA
targets

miRNA target
identification

False positives in
target prediction

Binding site
evolution after
miRNA loss

Strategy

Observed losses

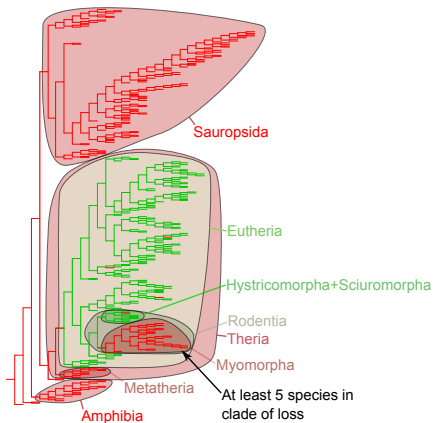
Evolution of seed match
count after loss

Conclusion

Supplementary
data

Observed losses

miR-656 family:



Overestimated
number of miRNA
targets

miRNA target
identification

False positives in
target prediction

Binding site
evolution after
miRNA loss

Strategy

Observed losses

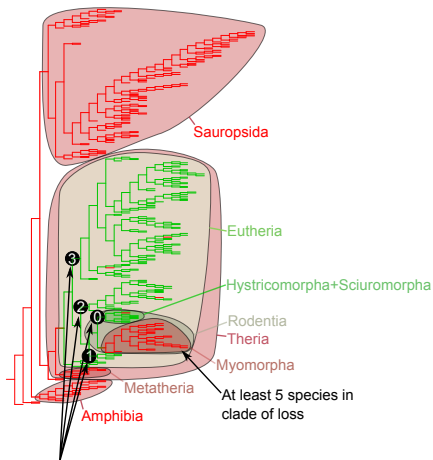
Evolution of seed match
count after loss

Conclusion

Supplementary
data

Observed losses

miR-656 family:



In sister clade + at least 3 degrees of cousinhood: miRNA family with certified presence in >50% of species at each degree

Overestimated number of miRNA targets

miRNA target identification

False positives in target prediction

Binding site evolution after miRNA loss

Strategy

Observed losses

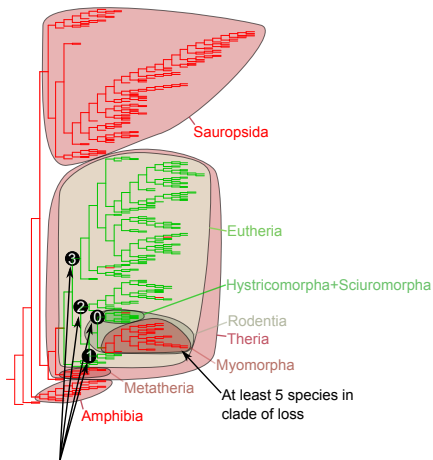
Evolution of seed match count after loss

Conclusion

Supplementary data

Observed losses

miR-656 family:



In sister clade + at least 3 degrees of cousinhood: miRNA family with certified presence in >50% of species at each degree

Overestimated number of miRNA targets

miRNA target identification

False positives in target prediction

Binding site evolution after miRNA loss

Strategy

Observed losses

Evolution of seed match count after loss

Conclusion

Supplementary data

Evolution of seed match count after loss

Overestimated
number of miRNA
targets

miRNA target
identification

False positives in
target prediction

Binding site
evolution after
miRNA loss

Strategy

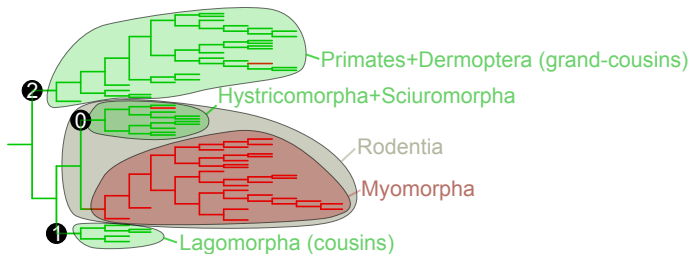
Observed losses

Evolution of seed match
count after loss

Conclusion

Supplementary
data

Evolution of seed match count after loss



Overestimated
number of miRNA
targets

miRNA target
identification

False positives in
target prediction

Binding site
evolution after
miRNA loss

Strategy

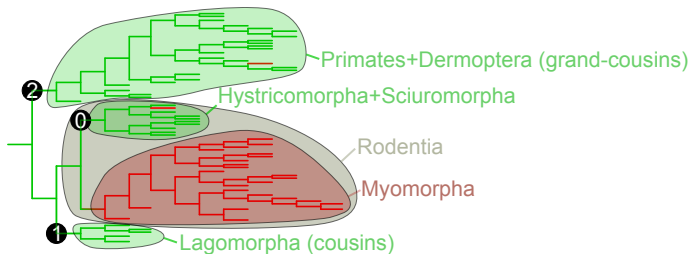
Observed losses

Evolution of seed match
count after loss

Conclusion

Supplementary
data

Evolution of seed match count after loss



PhyloP score: measures probability of selective conservation of seed match vs. rest of UTR.

Overestimated number of miRNA targets

miRNA target identification

False positives in target prediction

Binding site evolution after miRNA loss

Strategy

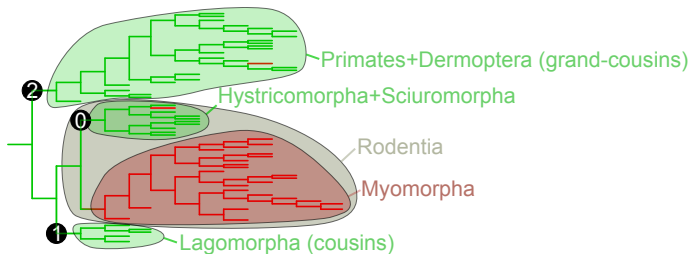
Observed losses

Evolution of seed match count after loss

Conclusion

Supplementary data

Evolution of seed match count after loss



PhyloP score: measures probability of selective conservation of seed match vs. rest of UTR in a clade vs. rest of tree.

Overestimated number of miRNA targets

miRNA target identification

False positives in target prediction

Binding site evolution after miRNA loss

Strategy

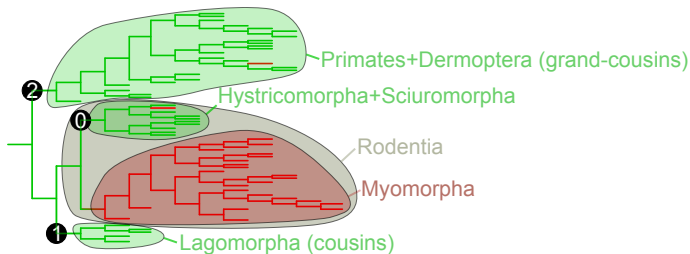
Observed losses

Evolution of seed match count after loss

Conclusion

Supplementary data

Evolution of seed match count after loss



PhyloP score: measures probability of selective conservation of seed match vs. rest of UTR in a clade vs. rest of tree.

→ Selects seed matches with selective conservation in rodents, then compare their conservation in rodent subclades (clade of loss vs. sister clade).

Overestimated number of miRNA targets

miRNA target identification

False positives in target prediction

Binding site evolution after miRNA loss

Strategy

Observed losses

Evolution of seed match count after loss

Conclusion

Supplementary data

Evolution of seed match count after loss

After miR-656 family loss:

Overestimated
number of miRNA
targets

miRNA target
identification

False positives in
target prediction

Binding site
evolution after
miRNA loss

Strategy

Observed losses

Evolution of seed match
count after loss

Conclusion

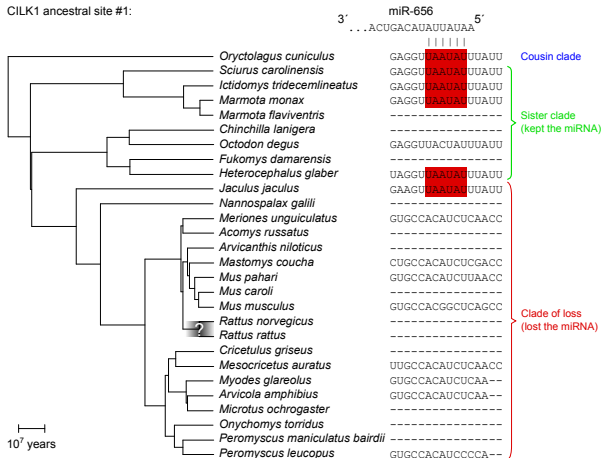
Supplementary
data

Evolution of seed match count after loss

Overestimated number of miRNA targets

After miR-656 family loss:

CILK1 ancestral site #1:



miRNA target identification

False positives in target prediction

Binding site evolution after miRNA loss

Strategy

Observed losses

Evolution of seed match count after loss

Conclusion

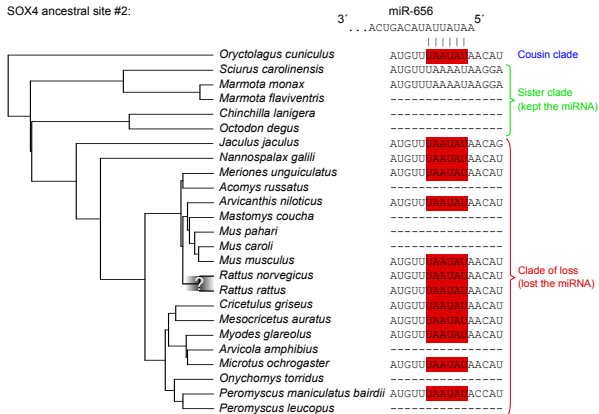
Supplementary data

Evolution of seed match count after loss

Overestimated
number of miRNA
targets

After miR-656 family loss:

SOX4 ancestral site #2:



miRNA target
identification

False positives in
target prediction

Binding site
evolution after
miRNA loss

Strategy

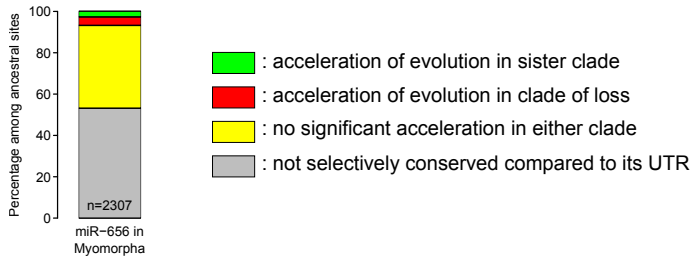
Observed losses

Evolution of seed match
count after loss

Conclusion

Supplementary
data

Evolution of seed match count after loss



Overestimated
number of miRNA
targets

miRNA target
identification

False positives in
target prediction

Binding site
evolution after
miRNA loss

Strategy

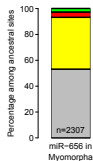
Observed losses

Evolution of seed match
count after loss

Conclusion

Supplementary
data

Evolution of seed match count after loss



- acceleration of evolution in sister clade
- acceleration of evolution in clade of loss
- no significant acceleration in either clade
- not selectively conserved compared to its UTR

Overestimated
number of miRNA
targets

miRNA target
identification

False positives in
target prediction

Binding site
evolution after
miRNA loss

Strategy

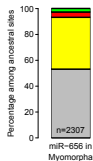
Observed losses

Evolution of seed match
count after loss

Conclusion

Supplementary
data

Evolution of seed match count after loss



- acceleration of evolution in sister clade
- acceleration of evolution in clade of loss
- no significant acceleration in either clade
- not selectively conserved compared to its UTR

Overestimated
number of miRNA
targets

miRNA target
identification

False positives in
target prediction

Binding site
evolution after
miRNA loss

Strategy

Observed losses

Evolution of seed match
count after loss

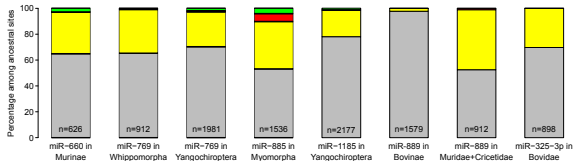
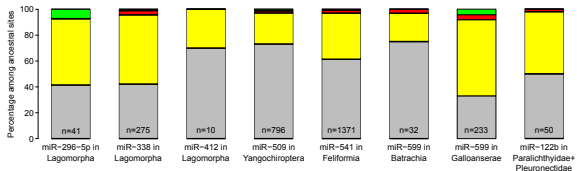
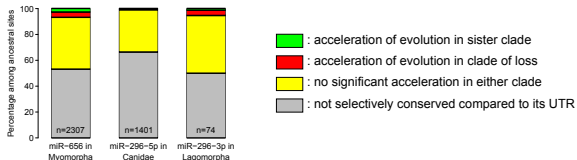
Conclusion

Supplementary
data

► Reasons for missing statistics

Evolution of seed match count after loss

Overestimated
number of miRNA
targets



miRNA target
identification

False positives in
target prediction

Binding site
evolution after
miRNA loss

Strategy

Observed losses

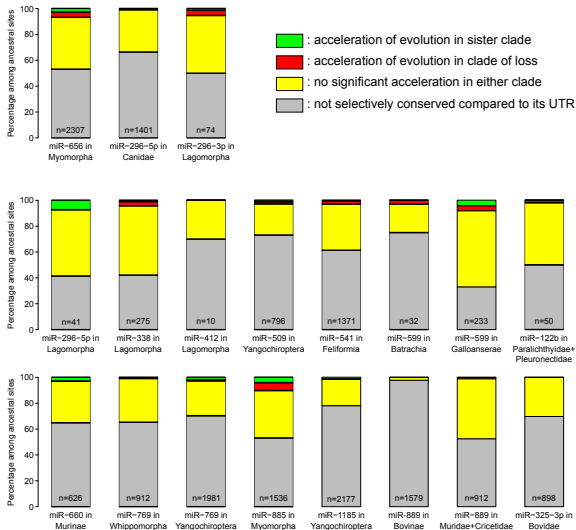
Evolution of seed match
count after loss

Conclusion

Supplementary
data

Evolution of seed match count after loss

Overestimated
number of miRNA
targets



miRNA target
identification

False positives in
target prediction

Binding site
evolution after
miRNA loss

Strategy

Observed losses

Evolution of seed match
count after loss

Conclusion

Supplementary
data

Most of these sites: conserved independently of the miRNA.

Conclusion

Overestimated
number of miRNA
targets

miRNA target
identification

False positives in
target prediction

Binding site
evolution after
miRNA loss

Strategy

Observed losses

Evolution of seed match
count after loss

Conclusion

Supplementary
data

Conclusion

High-throughput experimental target identification captures many false positives...

Overestimated
number of miRNA
targets

miRNA target
identification

False positives in
target prediction

Binding site
evolution after
miRNA loss

Strategy

Observed losses

Evolution of seed match
count after loss

Conclusion

Supplementary
data

Conclusion

High-throughput experimental target identification captures many false positives...

... but comparative genomics too!

Overestimated
number of miRNA
targets

miRNA target
identification

False positives in
target prediction

Binding site
evolution after
miRNA loss

Strategy

Observed losses

Evolution of seed match
count after loss

Conclusion

Supplementary
data

Conclusion

High-throughput experimental target identification captures many false positives...

... but comparative genomics too!

Improvement: identification of miRNA-independent conservation of binding sites.

Overestimated
number of miRNA
targets

miRNA target
identification

False positives in
target prediction

Binding site
evolution after
miRNA loss

Strategy

Observed losses

Evolution of seed match
count after loss

Conclusion

Supplementary
data

Conclusion

High-throughput experimental target identification captures many false positives...

... but comparative genomics too!

Improvement: identification of miRNA-independent conservation of binding sites.

On the analyzed examples: most conserved seed matches are not conserved because of the miRNA!

Overestimated
number of miRNA
targets

miRNA target
identification

False positives in
target prediction

Binding site
evolution after
miRNA loss

Strategy

Observed losses

Evolution of seed match
count after loss

Conclusion

Supplementary
data

Conclusion

High-throughput experimental target identification captures many false positives...

... but comparative genomics too!

Improvement: identification of miRNA-independent conservation of binding sites.

On the analyzed examples: most conserved seed matches are not conserved because of the miRNA!

→ A large part of the miRNA literature probably needs a cleanup.

Overestimated
number of miRNA
targets

miRNA target
identification

False positives in
target prediction

Binding site
evolution after
miRNA loss

Strategy

Observed losses

Evolution of seed match
count after loss

Conclusion

Supplementary
data

Conclusion

High-throughput experimental target identification captures many false positives...

... but comparative genomics too!

Improvement: identification of miRNA-independent conservation of binding sites.

On the analyzed examples: most conserved seed matches are not conserved because of the miRNA!

→ A large part of the miRNA literature probably needs a cleanup.

May be generalized to other regulators (transcription factors, RNA-binding proteins; [Seitz, 2024](#)).

Overestimated
number of miRNA
targets

miRNA target
identification

False positives in
target prediction

Binding site
evolution after
miRNA loss

Strategy

Observed losses

Evolution of seed match
count after loss

Conclusion

Supplementary
data

Acknowledgments



Current members:

Isabelle Busseau
Estelle Rebecq

Élisabeth Houbbron
Pauline De Sousa

Rémy Rodriguez

Former members:

Sophie Mockly
Natalia Pinzón
Laura Martinez

Germain Busto
Blaise Li
Anna Sergeeva

Overestimated
number of miRNA
targets

miRNA target
identification

False positives in
target prediction

Binding site
evolution after
miRNA loss

Strategy

Observed losses

Evolution of seed match
count after loss

Conclusion

Supplementary
data

Acknowledgments



Current members:

Isabelle Busseau
Estelle Rebecq

Élisabeth Houbbron
Pauline De Sousa

Rémy Rodriguez

Former members:

Sophie Mockly
Natalia Pinzón
Laura Martinez

Germain Busto
Blaise Li
Anna Sergeeva

Hiring for bioinfo projects!

Overestimated
number of miRNA
targets

miRNA target
identification

False positives in
target prediction

Binding site
evolution after
miRNA loss

Strategy

Observed losses

Evolution of seed match
count after loss

Conclusion

Supplementary
data

Acknowledgments



Current members:

Isabelle Busseau
Estelle Rebecq

Élisabeth Houbron
Pauline De Sousa

Rémy Rodriguez

Former members:

Sophie Mockly
Natalia Pinzón
Laura Martinez

Germain Busto
Blaise Li
Anna Sergeeva

Hiring for bioinfo projects!



Fondation
pour la recherche
sur le cancer



Overestimated
number of miRNA
targets

miRNA target
identification

False positives in
target prediction

Binding site
evolution after
miRNA loss

Strategy

Observed losses

Evolution of seed match
count after loss

Conclusion

Supplementary
data

Supplementary data

▶ *mir-656* orthologous loci in mammals

▶ Missing statistics

Overestimated
number of miRNA
targets

miRNA target
identification

False positives in
target prediction

Binding site
evolution after
miRNA loss

Strategy

Observed losses

Evolution of seed match
count after loss

Conclusion

Supplementary
data

Observed losses

miR-656 family:

Percentage identity among 470 mammals: 60 80 100



Overestimated
number of miRNA
targets

miRNA target
identification

False positives in
target prediction

Binding site
evolution after
miRNA loss

Strategy

Observed losses

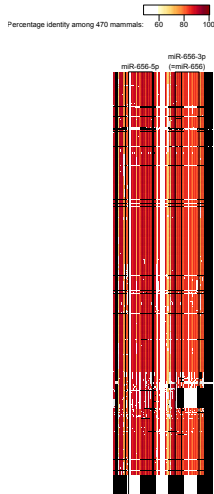
Evolution of seed match
count after loss

Conclusion

Supplementary
data

Observed losses

miR-656 family:



Overestimated number of miRNA targets

miRNA target identification

False positives in target prediction

Binding site evolution after miRNA loss

Strategy

Observed losses

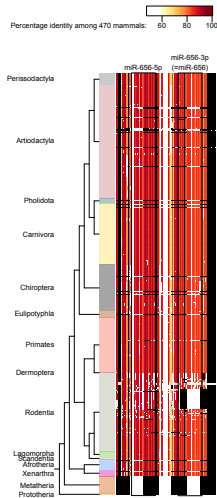
Evolution of seed match count after loss

Conclusion

Supplementary data

Observed losses

miR-656 family:



Overestimated
number of miRNA
targets

miRNA target
identification

False positives in
target prediction

Binding site
evolution after
miRNA loss

Strategy

Observed losses

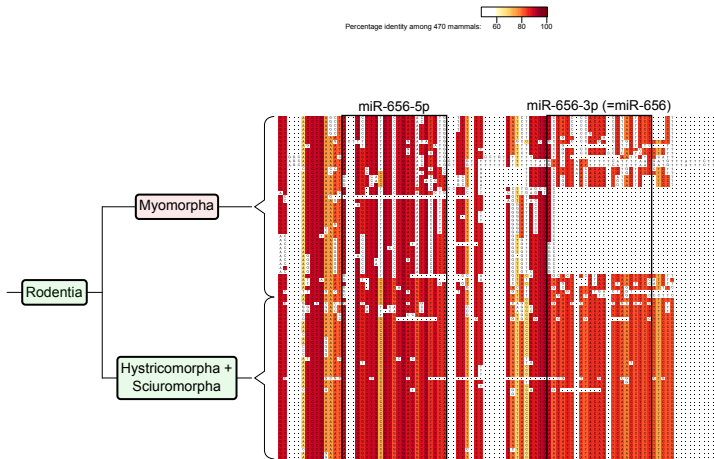
Evolution of seed match
count after loss

Conclusion

Supplementary
data

Observed losses

miR-656 family:



Overestimated number of miRNA targets

miRNA target identification

False positives in target prediction

Binding site evolution after miRNA loss

Strategy

Observed losses

Evolution of seed match count after loss

Conclusion

Supplementary data

Evolution of seed match count after loss

Overestimated
number of miRNA
targets

miRNA target
identification

False positives in
target prediction

Binding site
evolution after
miRNA loss

Strategy

Observed losses

Evolution of seed match
count after loss

Conclusion

Supplementary
data

Family	Seed	Clade of loss	Reason
miR-451a	AACCGU	Paralichthyidae+Pleuronectidae	<10 ancestral sites
miR-224	AAGUCA	Mustelidae	Under-studied sister clade
miR-338	ACAAUA	Syngnathidae	<10 ancestral sites
miR-460	ACAGCG	Serpentes	<10 ancestral sites
miR-880	ACUCCA	Deomyiinae+Gerbilinae	Under-studied sister clade
miR-452	ACUGUU	Mustelidae	Under-studied sister clade
miR-456	AGGCUG	Balistidae+Tetraodontidae	Under-studied sister clade
miR-2188	AGGUCC	Eutheria	<10 ancestral sites
miR-2114	AGUCCC	Phocidae	Under-studied sister clade
miR-654	AUGUCU	Delphinoida	Under-studied sister clade
miR-460b	CCUCAU	Mammalia	<10 ancestral sites
miR-460b	CCUCAU	Serpentes	<10 ancestral sites
miR-676	CGUCCU	Cervidae+Bovidae	Under-studied sister clade
miR-676	CGUCCU	Feloidea	<10 ancestral sites
miR-187	CGUGUC	Serpentes	<10 ancestral sites
miR-509	GAUUGA	Cervidae+Bovidae	Under-studied sister clade
miR-31	GGCAAG	Protacanthopterygii	<10 ancestral sites
miR-31	GGCAAG	Lophiiformes+Tetraodontiformes	<10 ancestral sites
miR-458	UAGCUC	Theria	<10 ancestral sites
miR-374	UAUAAU	Yangochiroptera	In progress
miR-513b	UCACAA	Bos	<10 ancestral sites
miR-653	UGAAAC	Yangochiroptera	In progress
miR-1662	UGACAU	Endoglyptodonta	Under-studied sister clade
miR-297p	UGUAUG	Pinnipedia+Musteloidea	Under-studied sister clade
miR-2131	UGUUAC	Psittaciformes	Under-studied sister clade
miR-599	UUGAUA	Corvides+Passerides	Under-studied sister clade