Evolution of glyoxylate cycle-specific enzymes

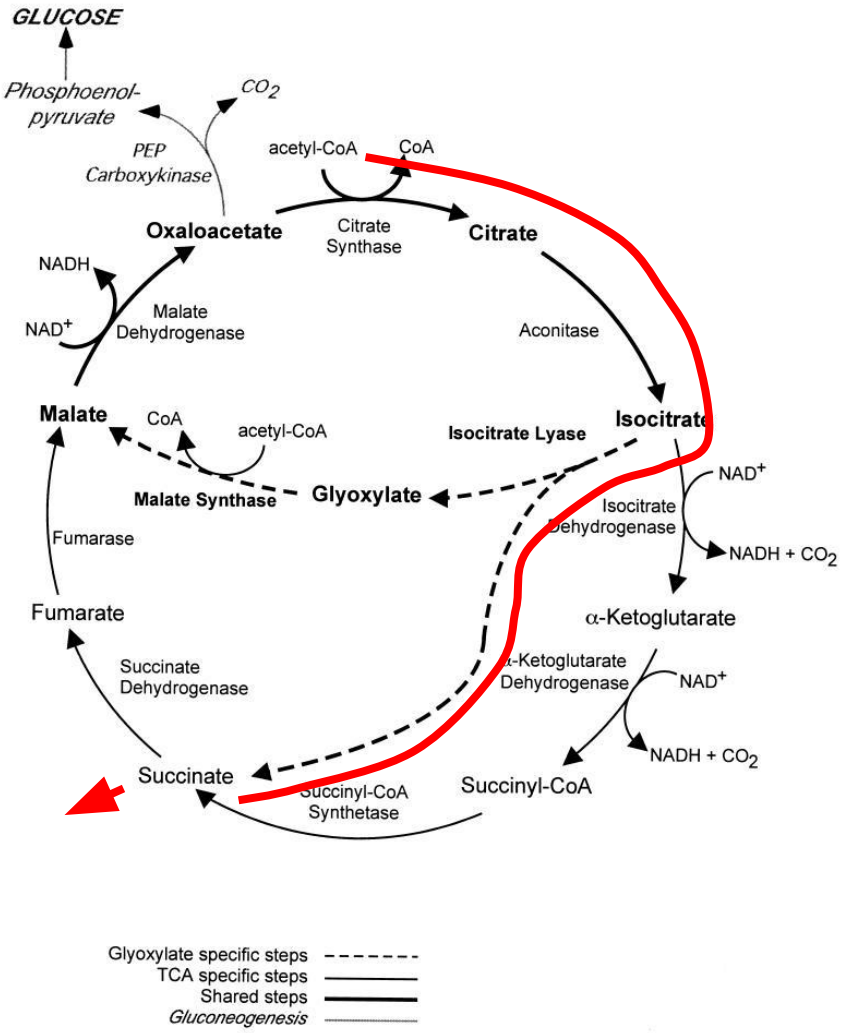
Sofia Bichurina, Anastasiya Golubeva, Petr Ekaterinochkin, Ekaterina Ignatova, Ivan Kozin, Maria Kulbachnaya, Nikita Litvyakov, Tatiana Murakhovskaya, Ekaterina Maksimova, Fyodor Kondrashov

Glyoxylate cycle

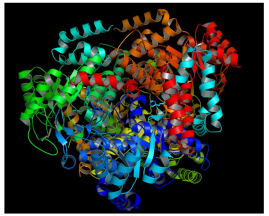
Is an anabolic pathway that is used to synthesize succinate from acetyl-CoA.

Thanks to GS acetate can be used as a precursor for synthesis of various organic compounds.

For example, it is used in lipid-rich seeds to initiate growth at a time when the plant does not photosynthesize, so the lipids can be turned to sugars.



Isocitrate lyase (ICL)



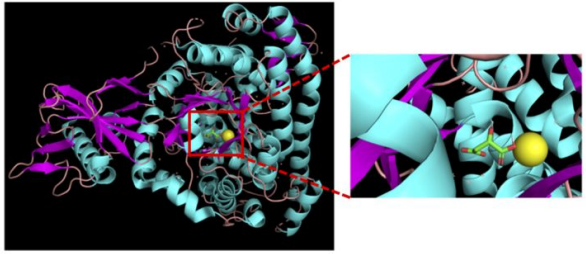
The glyoxylate cycle has two specific enzymes, Isocitrate lyase and malate synthase.

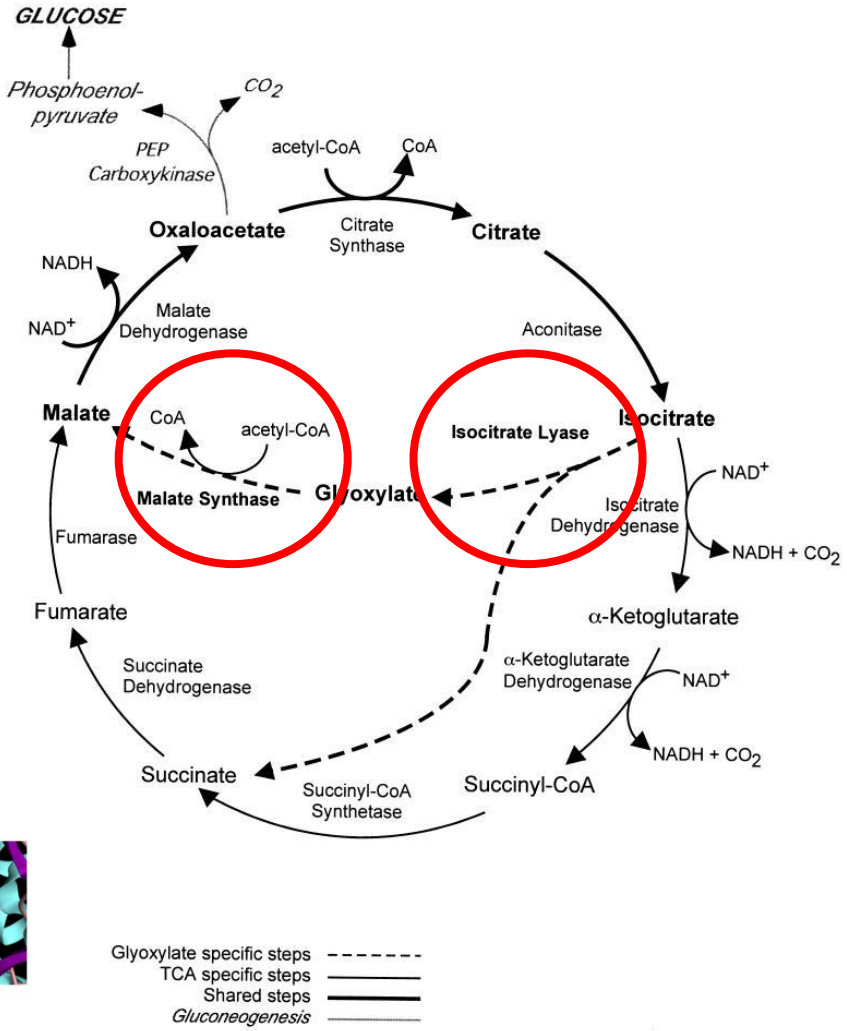
Isocitrate lyase turns isocitrate into succinate and glyoxylate .

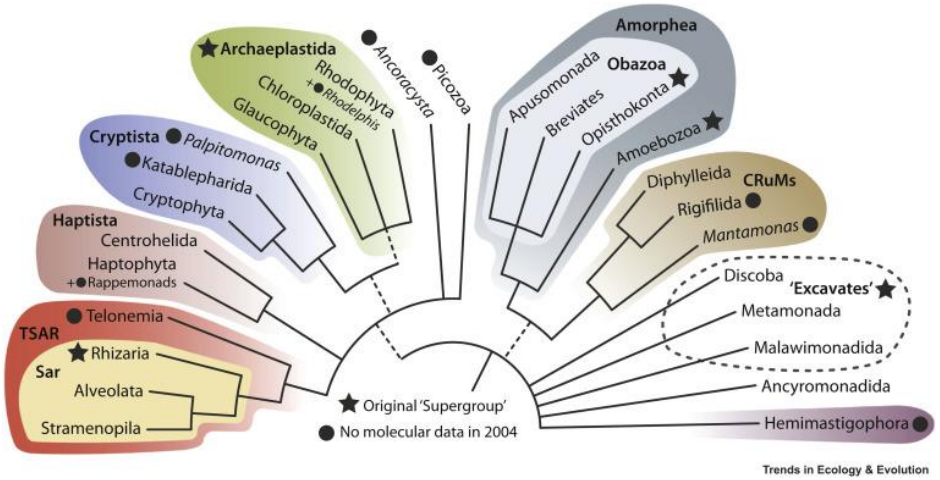
Malate synthase turns glyozylate and acetyl-CoA into malate.

To study the evolution of the glyoxylate cycle it makes sense to focus on these two key enzymes.

Malate synthase (MS)

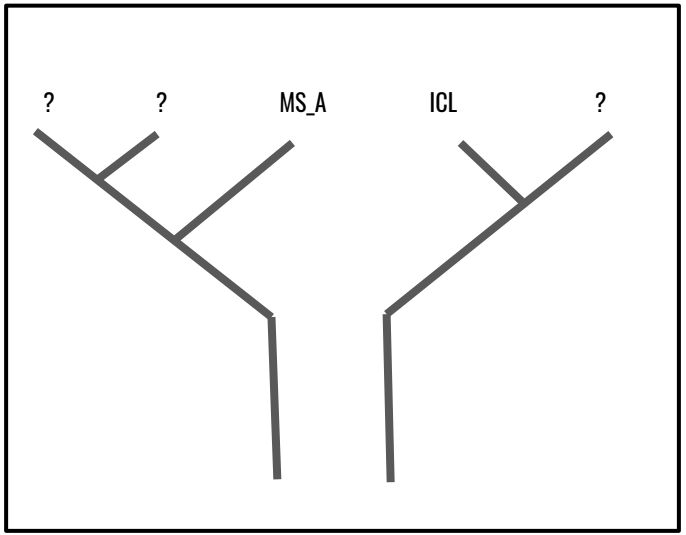






Previous work has shown that the glyoxylate cycle-specific enzymes may have been horizontally transferred into eukaryotes.

Our work focused on looking for evidence of horizontal gene transfer of ICL and MS using a much larger dataset of these two enzymes from completely sequenced genomes..



Workflow

The trees indicated which sequences were more closely related.

Tree topology

Conclusions

Our results were used to reconstruct the evolutionary trajectories of the glyoxylate cycle-specific enzymes, how they were lost or gained.

We analyzed the topology of the phylogenetic trees.

Analysis

Trees

The alignments were used to build trees.

We were using previously described sequences to find new ones.

NCBI BLAST

Data

The found sequences were aligned.

BLAST allowed us to find orthologous sequences..

Methods

We used **BLAST** to search for ILC & MS in: Bacteria, Archaea and Eukaryotes. Search was recursive, starting from E. coli (IC1L: MHO05562 MSa: MHO05561).

Muscle as implemented in MEGA:

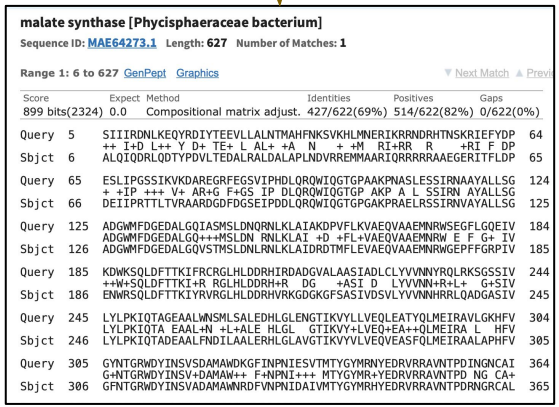
Alignment and building the 1st phylogenetic tree by Maximum Likelihood method.

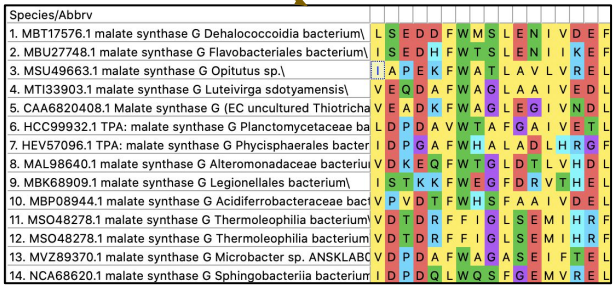
MrBayes:

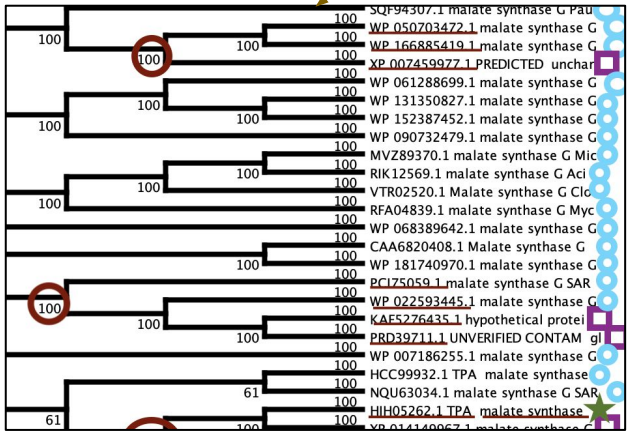
The second tree-construction in command line interface with default settings.

Mesquite:

Data visualization from MrBayes in a more convenient way.







ICL and MS that were found

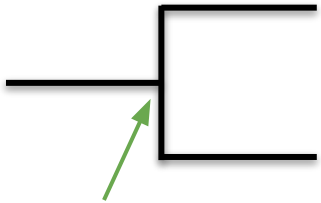
	ICL1	ICLz	ICLh	MSa	MSg	MSi	MSh	MSd
Eukaryotes	+	+	-	+	+	+	-	-
Bacteria	+	-	-	+	-	-	-	+
Archaea	-	-	+	-	-	-	+	-

1. ICLz, MSg and MSi were rarely found in eukaryotes. ICL1 and MSa were much more frequent.

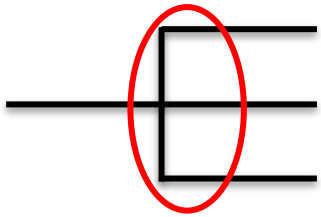
2.Many eukaryotes did not have both enzymes.

3.Enzyme ICL1 and MSa were frequently found in bacteria.

4. Archaea rarely had these enzymes.

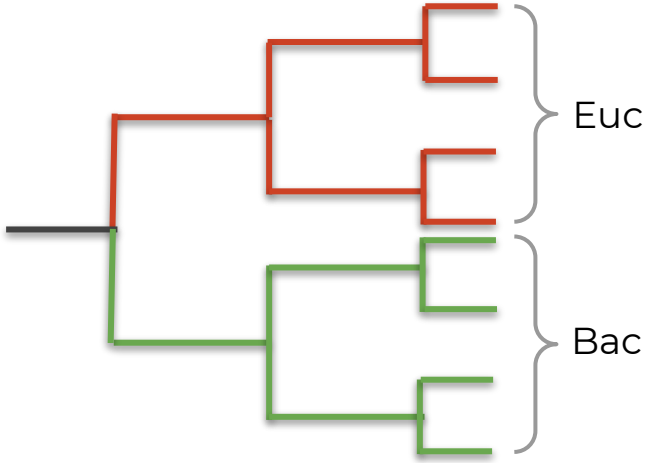


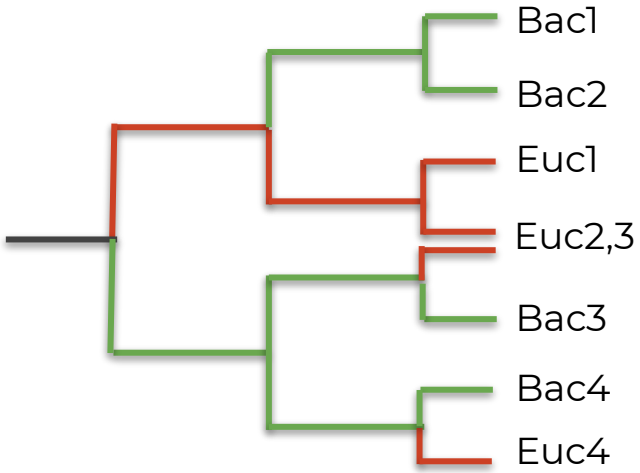
Evolutionary event / occasion



Unresolved phylogenetic tree

- Each branch on a phylogenetic tree represents only one evolutionary event.
- The simultaneous existence of two or more evolutionary events is impossible.
- Usually, proteins of organisms of different groups are located separately from each other on a phylogenetic tree.





Contamination

VS

Horizontal gene transfer

Contamination

Contamination can occur during sequencing. For instance, sometimes bacterial DNA fragment can be found in a sequence of the sample organism.

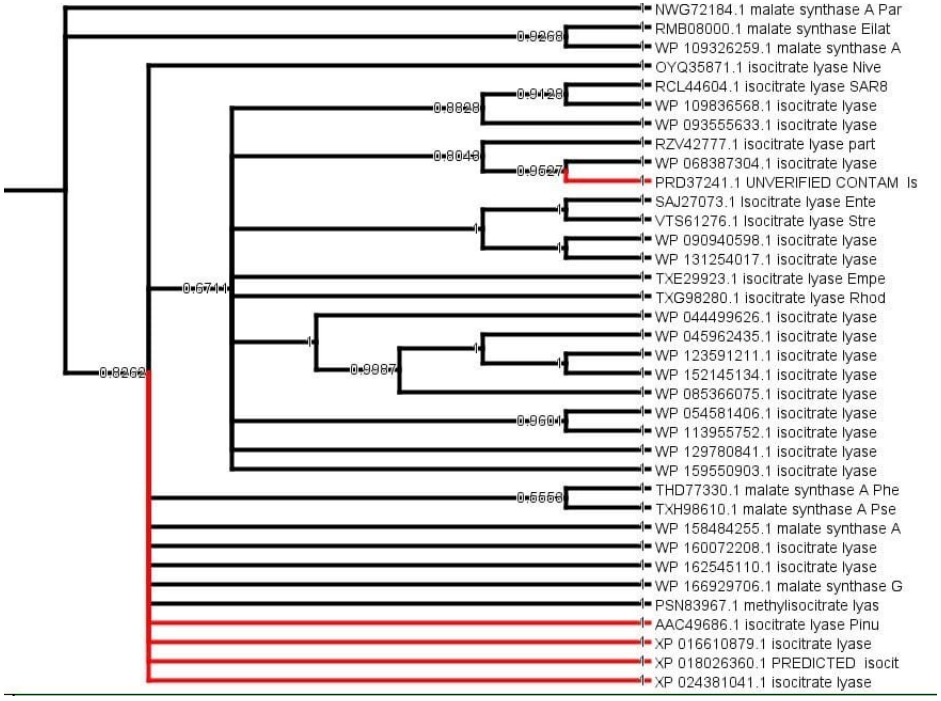
We battled with it during BLAST-stage, excluding suspicious sequences, such as:

- From incompleted genomes
- Short and poor alignments
- Weirdly perfect alignments

We used reverse-BLAST & pairwise comparison methods to distinguish contamination. Yet some contaminations likely got into the final chart.

- Bacteria
- Eukaryotes

Example of contamination



In this tree you can see that most eukaryotes are separated from bacteria, but one organism is near.. Looks like horizontal gene transfer.

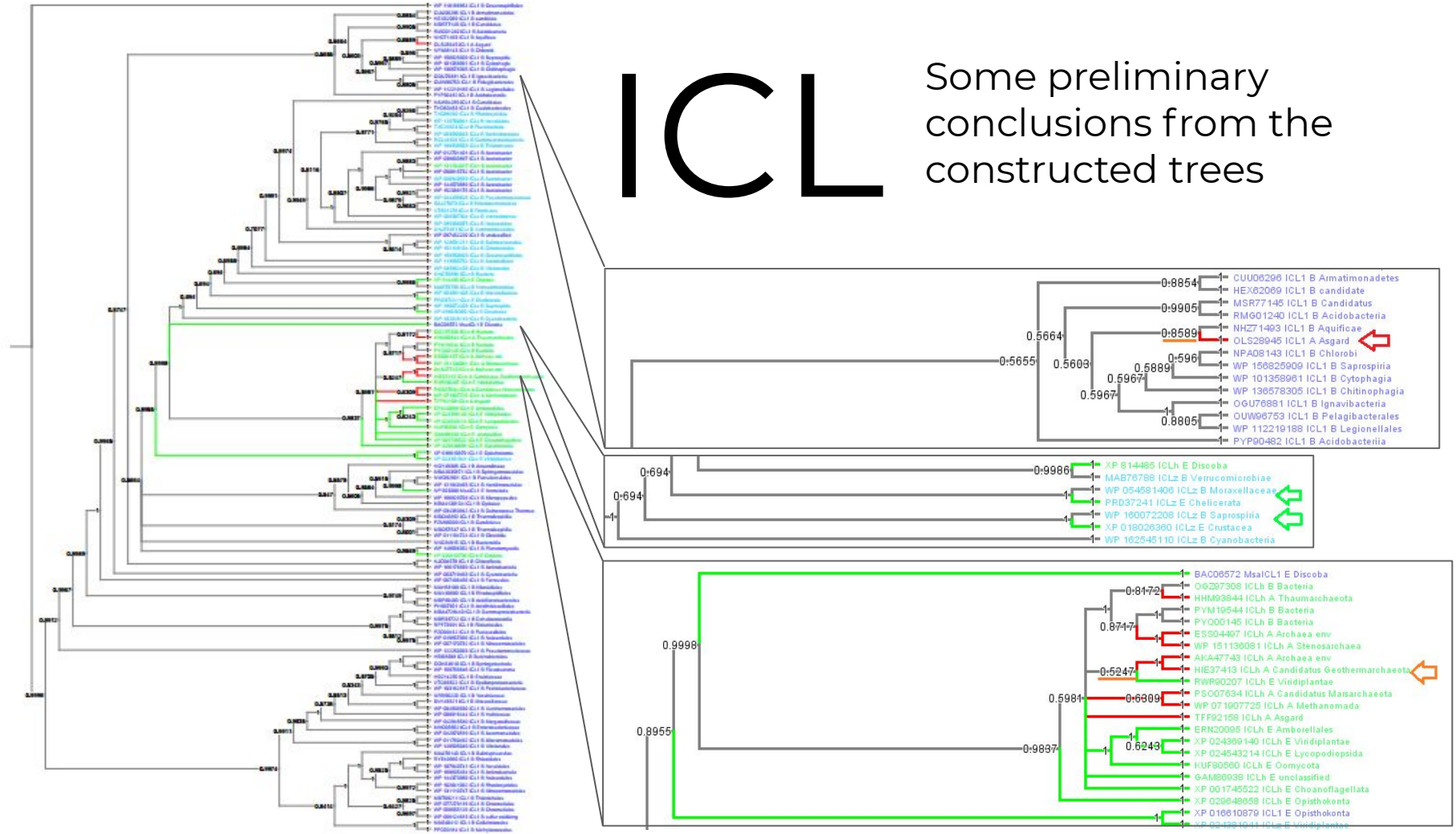
But checking pairwise alignments confirms that it is contamination - the sequences are just too similar. Thus, no horizontal gene transfer was found.

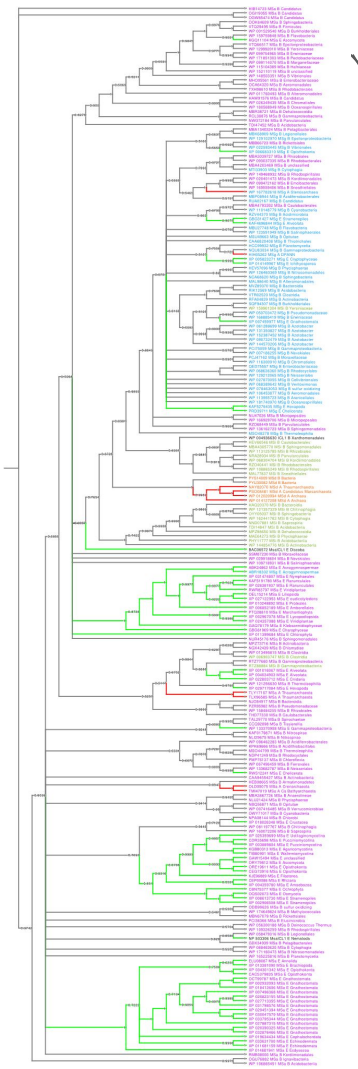
Score	Expect	Method	Identities	Positives	Gaps
768 bits(1982)	0.0	Compositional matrix adjust.	368/524(70%)	429/524(81%)	3/524(0%)

```
Query  9    HVRAVKEKLGGT--WGAIRPEDAARMMVQNRFHTGLDIAKYTAAIMRKDMAEYDADSSKY  66
            H++A+ E  G    W AI PE AARM +QNRF +GLDIA+YTAAIMR+DMAEYD D+S+Y
Sbjct  7    HIKALTELAGKNPNWAAINPEYAARMRLQNRFKSGLDIARYTAAIMRRDMAEYDQDASRY  66

Query  67   TQSLGCWHGFIAQQKMIANKKYFGTTSKRYIYLSGWMVAALRSEFGPLPDQSMHEKTSVP  126
            TQSLGCWHGFIAQQK+IA KK+  +T KRY+YLSGWM+AALRSEFGPLPDQSMHEKT+VP
Sbjct  67   TQSLGCWHGFIAQQKLIAIKKHQKSTEKRYLYLSGWMIAALRSEFGPLPDQSMHEKTAVP  126

Query  127  KLIAEIYTFLRQADAKELNDLFRALQKAEQAGDSAKAAEIISQIDNFESHVVPIIADIDA  186
             LI EIY FLRQAD +EL+ LF AL KA  AG++ +   +++QI+ F++HVVPIIADIDA
Sbjct  127  ALIGEIYDFLRQADTRELDLLFSALDKARAAGETREETALLAQIEGFQTHVVPIIADIDA  186

Query  187  GFGNEEATYLLTKQMIEAGACAIQIENQVSDAKQCGHQAGKVTVPHEDFLAKINAVRYAF  246
            GFGN EATYLL K+MIEAGAC IQIENQVSD KQCGHQ GKVTVPH DFLAKINAVRYAF
Sbjct  187  GFGNAEATYLLAKKMIEAGACCIQIENQVSDEKQCGHQDGKVTVPHADFLAKINAVRYAF  246

Query  247  LELGVDEGVIVARTDSEGADLTQKIPVSKEKGDLASQYISYLDTKEIDISEASDDEILIK  306
            LELGVD+GVIVARTDS GA LT++I VS++ GDL  QY  +LD  EI  +E  + E+LI
Sbjct  247  LELGVDDGVIVARTDSLGAGLTKQIAVSEKPGDLGDQYNRFLDCDEIHPAELQNGEVLIS  306

Query  307  RDGKLHRPTRLASGLYQFREGTQHDRVVLDCVTSLQNGADMIWIETPTPDVAGIAGFVND  366
            RDGKL RP RLASGL+QF++G+  DR VLDC+TSLQNGADMIWIET  P +  I G ++
Sbjct  307  RDGKLLRPKRLASGLFQFQKGSGEDRCVLDCITSLQNGADMIWIETEKPHIGQIKGMLDR  366

Query  367  IKKQVPNAKLVYNNSPSFNWTLNFRQQAYDRWVAEGKDVSGYDRAKLMSAEYDNSELAAD  426
            I++ VPNAKLVYNNSPSFNWTLNFRQQ +D W AEG+D++ Y R  LMSAEYD +EL
Sbjct  367  IREVVPNAKLVYNNSPSFNWTLNFRQQVFDAWQAEGRDLAAYQRDALMSAEYDETELGKV  426

Query  427  ADEKIRTFQADAAREAGVFHHLITLPTYHTAALSTHELAKGYFGEEGMLAYVAGVQRKEI  486
            AD KI++FQ DA+R+AGVFHHLITLPTYHTAALST  L++GYFG + MLAYV GVQR+EI
Sbjct  427  ADAKIQSFQRDASRDAGVFHHLITLPTYHTAALSTDMLSRGYFGSDAMLAYVKGVQRQEI  486

Query  487  RGGIACVKHQAMAGSDIGDDHKEIFAGENALKAGDASKNTMNQF  530
            R GIACVKHQ MAGSDIGD+HKE FAG  ALKA     NTM+QF
Sbjct  487  RRGIACVKHQNMAGSDIGDNHKEYFAGVAALKAS-GKDNTMHQF  529
```

ICL

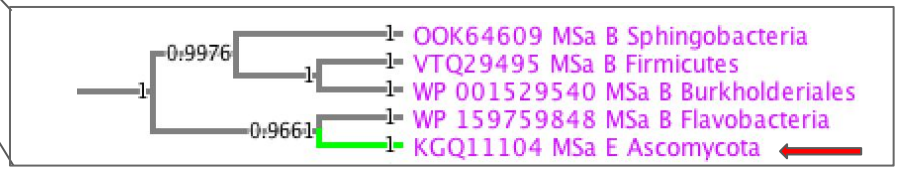
Some preliminary conclusions from the constructed trees

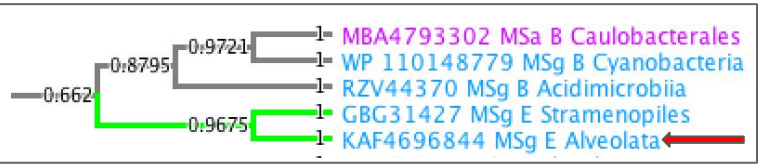


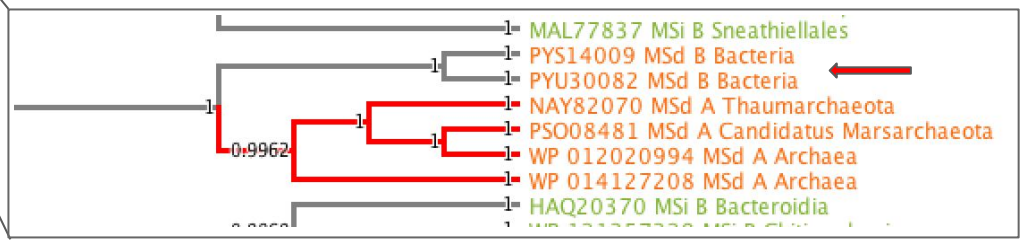


MS

Some preliminary conclusions from the constructed trees







Conclusion

1. We found two new forms ICL and three new forms MSa
2. We found that most of eukaryotic enzymes glyacid cycle clusterize with archaeal. In indicates common ancestor and subsequent loss in many clades of eukaryotes.
3. Some cases are similar to both horizontal transfer and contamination.
4. Sometimes existence glyacid cycle not proven to contain enzymes

What else can you work on?

1. Increasing the accuracy of tree alignment.
2. Prove existence glyoxylate cycle experimentally and reads articles about studied cases.
3. Study the patterns of the combination of isoforms in different organisms.
4. Correct results after appearance new data about living organisms.