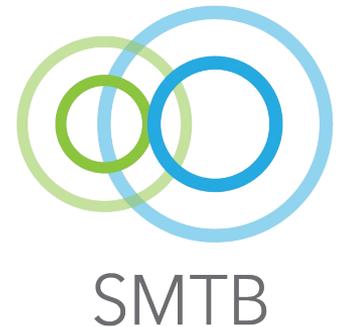


# Evolution of glyoxylate cycle-specific enzymes

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

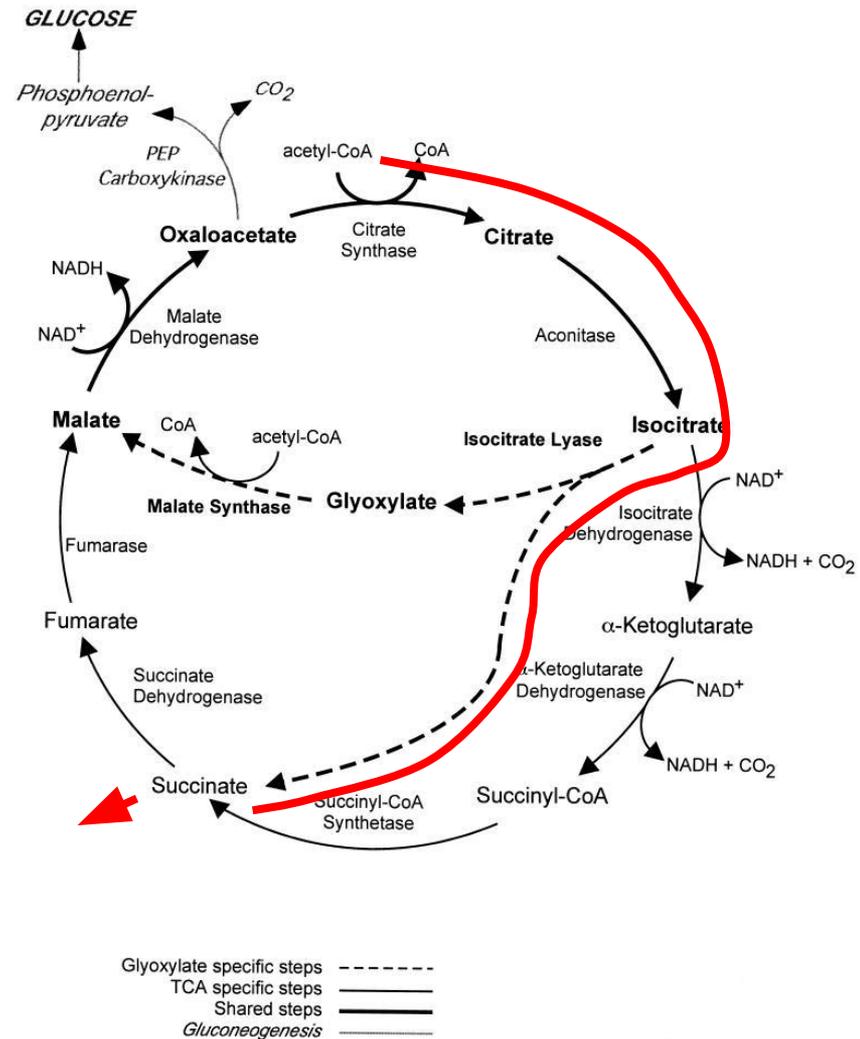


# Glyoxylate cycle

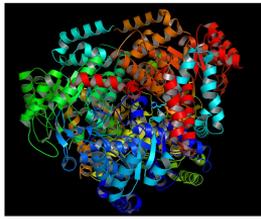
Is an anabolic pathway that is used to synthesize succinate from acetyl-CoA.

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)



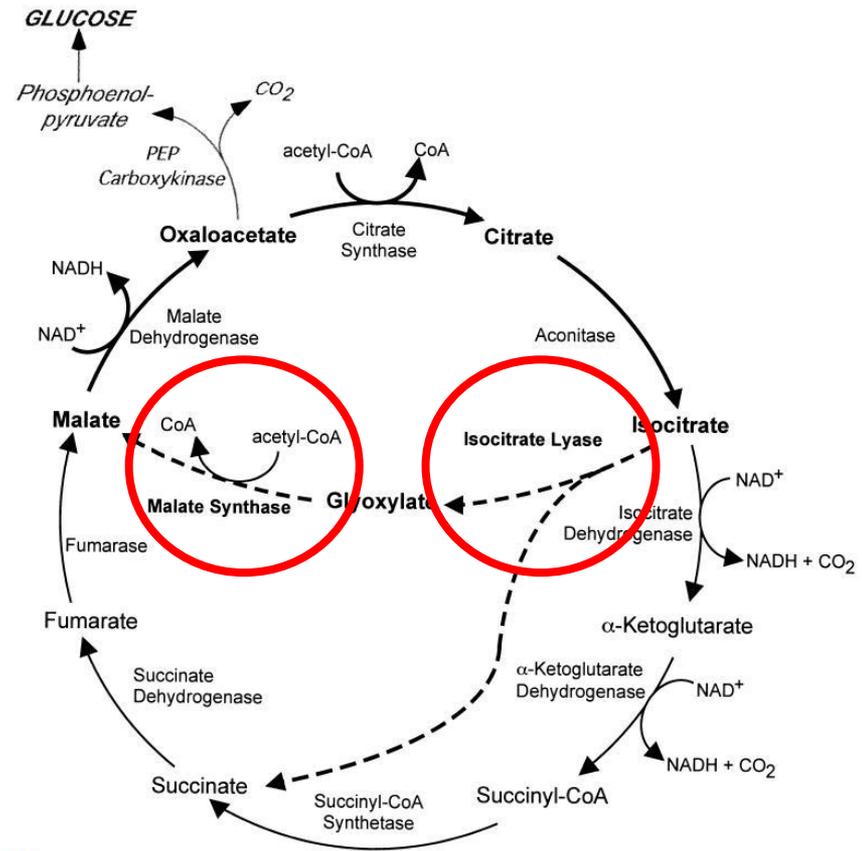
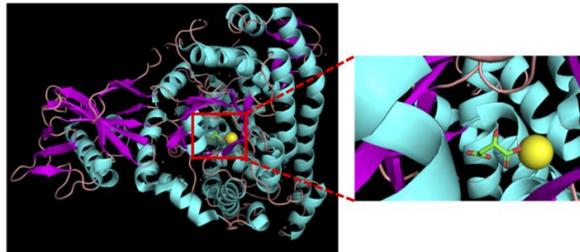
The glyoxylate cycle has two specific enzymes, Isocitrate lyase and malate synthase.

Isocitrate lyase turns isocitrate into succinate and glyoxylate.

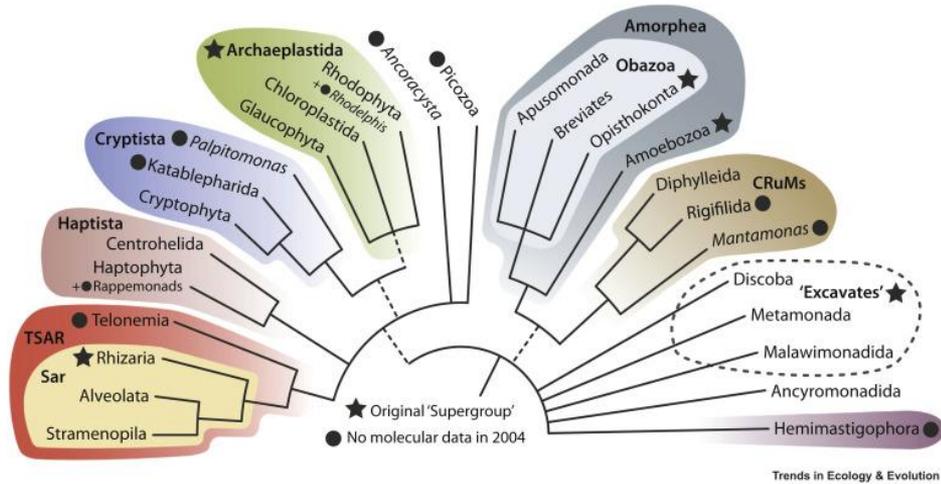
Malate synthase turns glyoxylate and acetyl-CoA into malate.

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

## Malate synthase (MS)

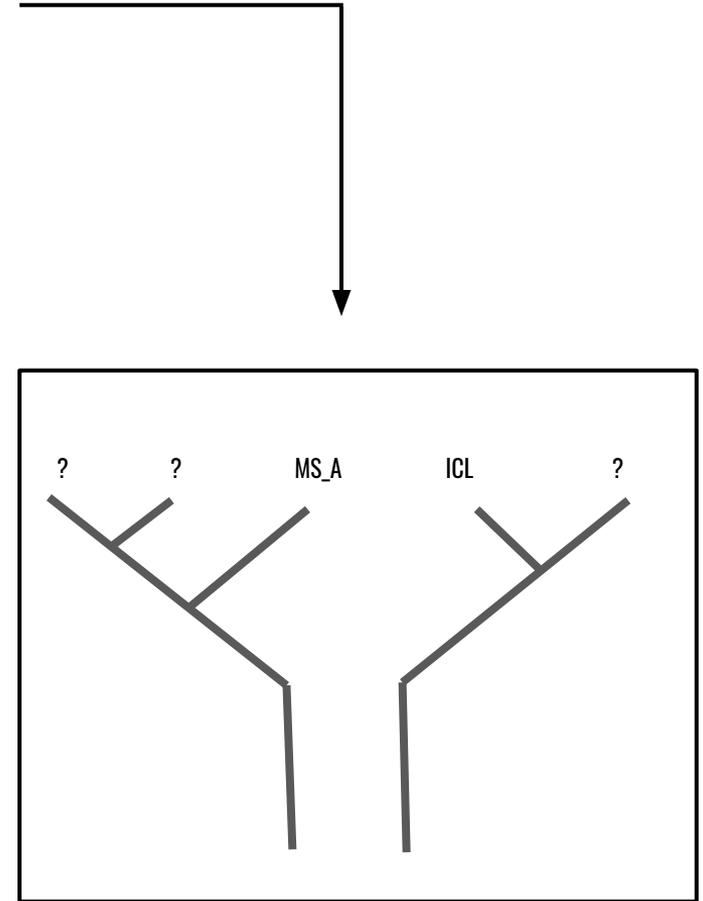


Glyoxylate specific steps -----  
 TCA specific steps =====  
 Shared steps =====  
 Gluconeogenesis =====

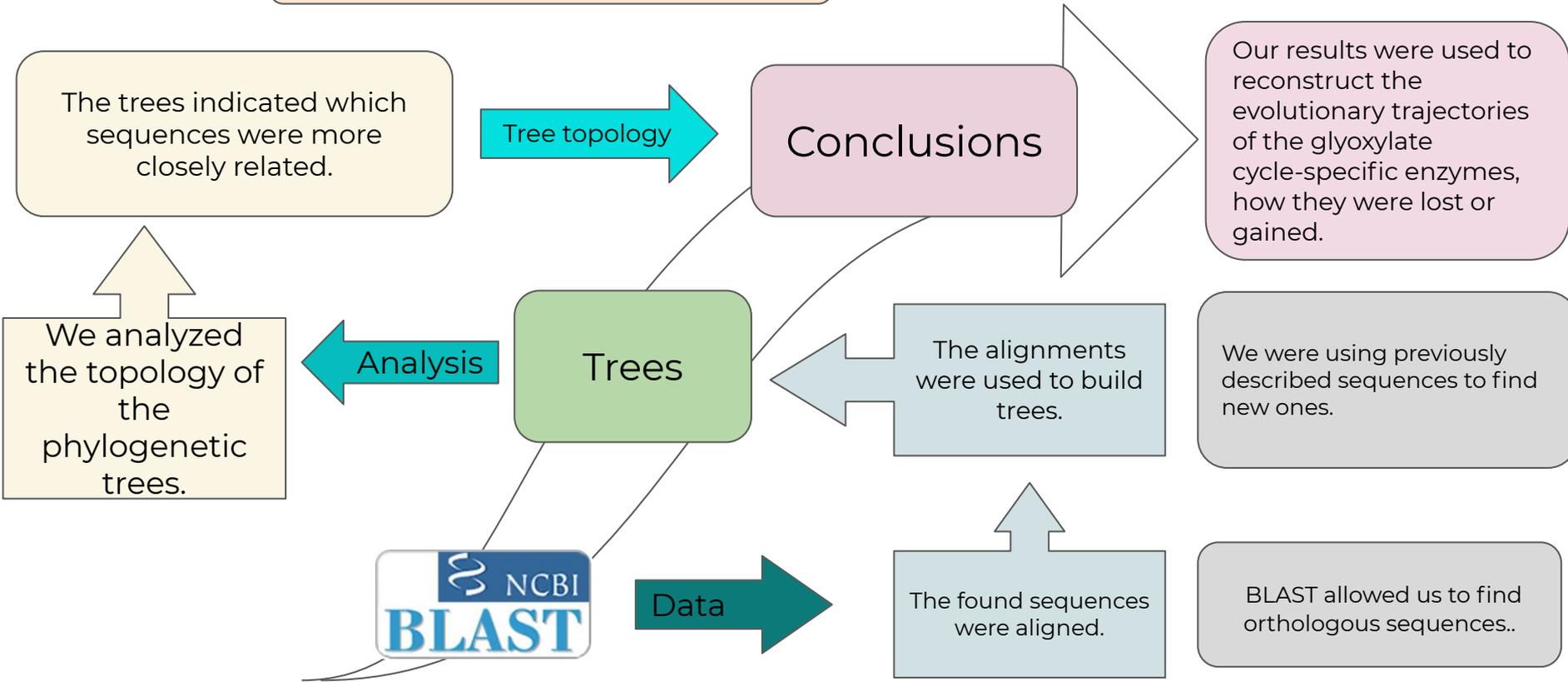


Previous work has shown that the glyoxylate cycle-specific enzymes may have been horizontally transferred into eukaryotes.

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



# Methods

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

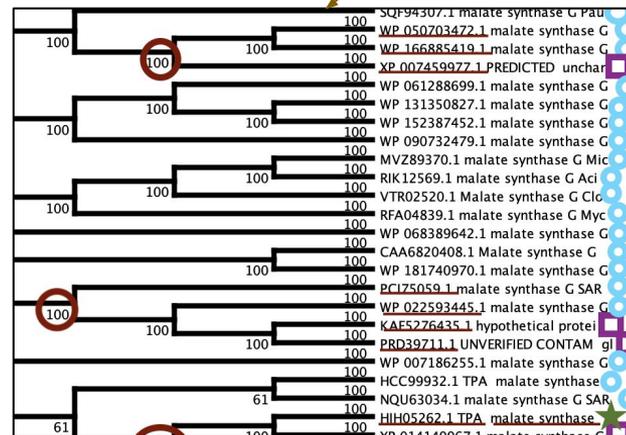
## malate synthase [Phycisphaeraeae bacterium]

Sequence ID: MAE64273.1 Length: 627 Number of Matches: 1

Range 1: 6 to 627 GenPept Graphics

Score	Expect	Method	Identities	Positives	Gaps
899 bits(2324)	0.0	Compositional matrix adjust.	427/622(69%)	514/622(82%)	0/622(0%)
Query 5	SIIRDNLKEQRYDIYEEVLLALNTMAHFNKSVKHLMNERIKRRNDRHTNSKRIFEYDP	64			
Sbjct 6	ALQIQDLRLQDYDPVLTEDLRALDALAPLNDVRRREMAARIQRRRRRAAEEGERITFLDP	65			
Query 65	ESLTPGSSIKVDAREGRFEGSVIPHLQRQWIGTQPAKPNASLESIRNAAYALLSG	124			
Sbjct 66	DEIIPRITTLTVRAARDGDFGSEIPDLQRQWIGTQPAKAPALSSIRNAYALLSG	125			
Query 125	ADGWMFDGEDALGQVSTMSLDNRLNKLAIADKPVFLKVAEQVAEMNRWSEFGLQEIY	184			
Sbjct 126	ADGWMFDGEDALGQVSTMSLDNRLNKLAIADKPVFLKVAEQVAEMNRWSEFGLQEIY	185			
Query 185	KDWKSQLDFTTKIFRCRGLHDDRHRDADVALAASIDLQVYVNNYRQLRKSQSSIV	244			
Sbjct 186	ENWRSQDLFTTKIYRVRGLHDDRHRGDKGKFSASIVDSLVYVNNHRRLOADGASIV	245			
Query 245	LYLPKIQTAGEAALWNSMLSALEDHGLGENTIKYVLLVEQLYATYQMEIRAVLQKHFV	304			
Sbjct 246	LYLPKIQTAEAAFLNDILAERHLGLAVGTIKYVLLVEQVEASQFMEIRAALAPHFV	305			
Query 305	GYNTGRWDYINSVDAMAWDKGFINPNIIESVTMTYGYMRNYEDRVRRAVNTPDINGCAI	364			
Sbjct 306	GFNTGRWDYINSVDAMAWNRDFVNPNIIDAVMTYGYMRHYEDRVRRAVNTPDNRGRCAL	365			

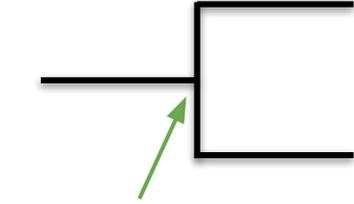
Species/Abbrv	L	S	E	D	D	F	W	M	S	L	E	N	I	V	D	E	F
1. MBT17576.1 malate synthase G Dehalococcoidia bacterium\	L	S	E	D	D	F	W	M	S	L	E	N	I	V	D	E	F
2. MBU27748.1 malate synthase G Flavobacteriales bacterium\	L	S	E	D	H	F	W	T	S	L	E	N	I	K	E	F	
3. MSU49663.1 malate synthase G Opitutus sp.)	A	P	E	K	F	W	A	T	L	A	V	L	V	R	E	L	
4. MTI33903.1 malate synthase G Luteivirga sdotyamensis\	V	E	Q	D	A	F	W	A	G	L	A	A	I	V	E	D	L
5. CAA6820408.1 Malate synthase G (EC uncultured Thiotricha	V	E	A	D	K	F	W	A	G	L	E	G	I	V	N	D	L
6. HCC99932.1 TPA: malate synthase G Planctomycetaceae ba	L	D	P	A	V	T	A	F	G	A	I	V	E	T	L		
7. HEV57096.1 TPA: malate synthase G Phycisphaerales bacter	L	D	P	A	F	W	H	A	L	A	D	L	H	R	G	F	
8. MAL98640.1 malate synthase G Alteromonadaceae bacterium\	V	D	K	E	Q	F	W	T	G	L	D	T	L	V	H	D	
9. MBK68909.1 malate synthase G Legionellales bacterium\	I	S	T	K	K	F	W	E	G	F	D	R	V	T	H	E	L
10. MBP08944.1 malate synthase G Acidiferrobacteraceae bac	V	P	D	T	F	W	H	S	F	A	A	I	V	E	D	L	
11. MSO48278.1 malate synthase G Thermoleophilla bacterium\	V	D	T	R	F	F	I	G	L	S	E	M	I	H	R	F	
12. MSO48278.1 malate synthase G Thermoleophilla bacterium\	V	D	T	R	F	F	I	G	L	S	E	M	I	H	R	F	
13. MVZ89370.1 malate synthase G Microbacter sp. ANSKLABC	V	D	P	A	F	W	A	G	A	S	E	I	F	T	E	L	
14. NCA68620.1 malate synthase G Sphingobacteriia bacterium\	I	D	P	Q	L	W	Q	S	F	G	E	M	V	R	E	L	



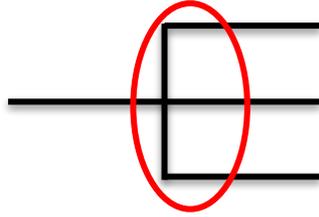
# ICL and MS that were found

	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.

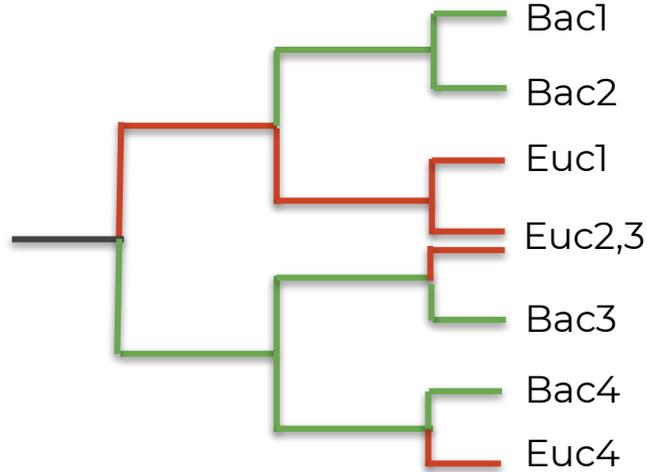
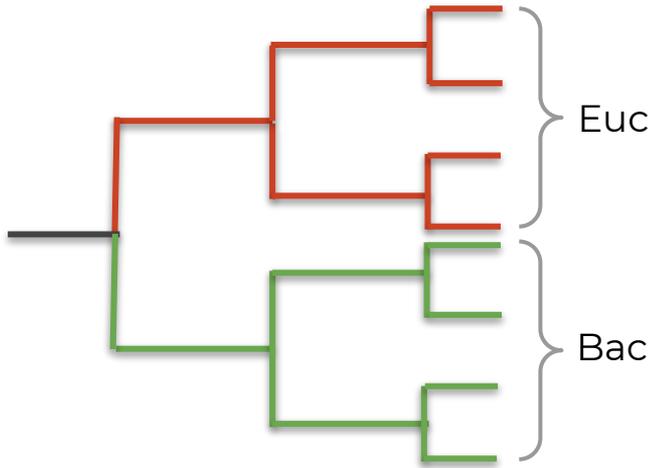


Evolutionary event / occasion



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.

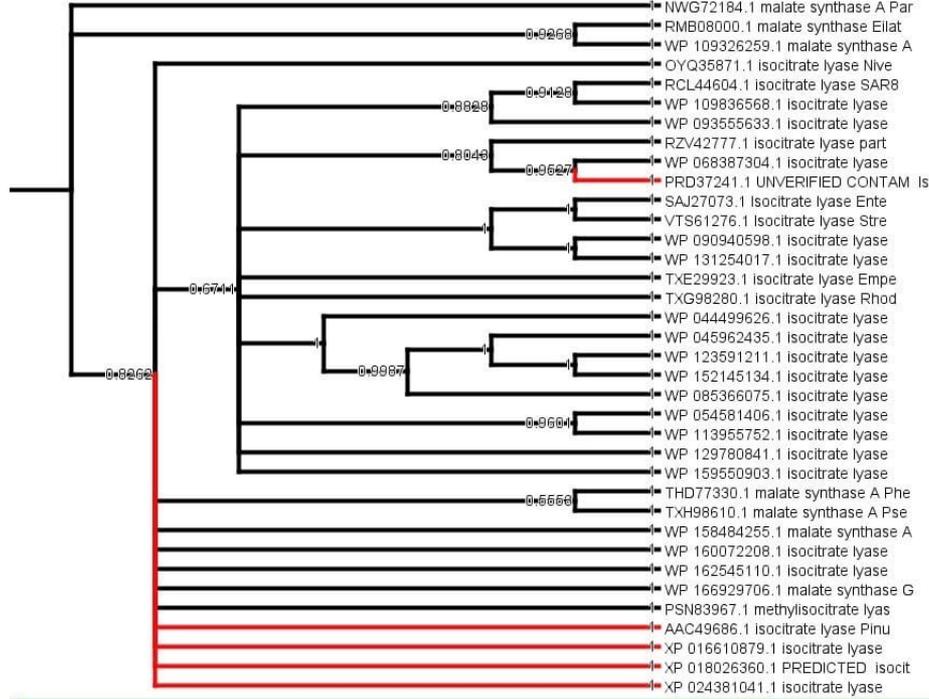
We battled with it during BLAST-stage, excluding suspicious sequences, such as:

- From incompleting 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



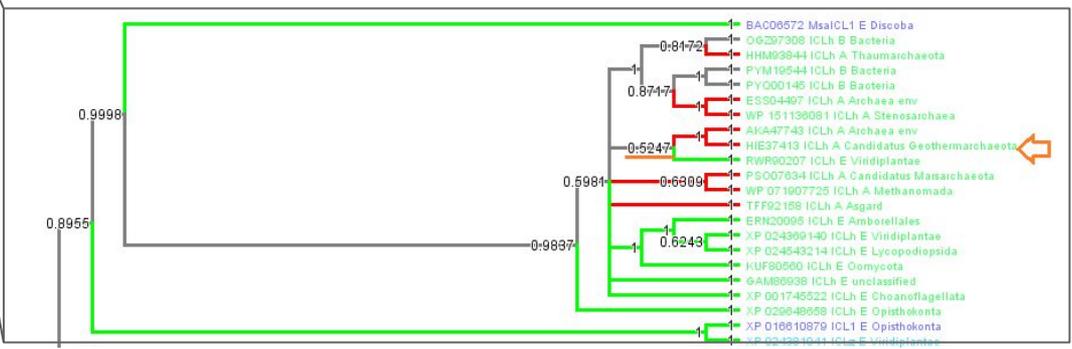
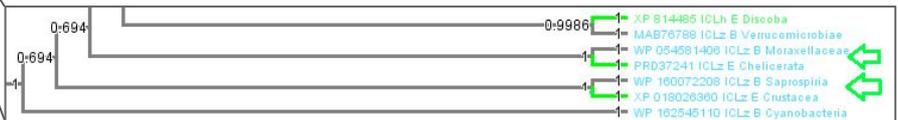
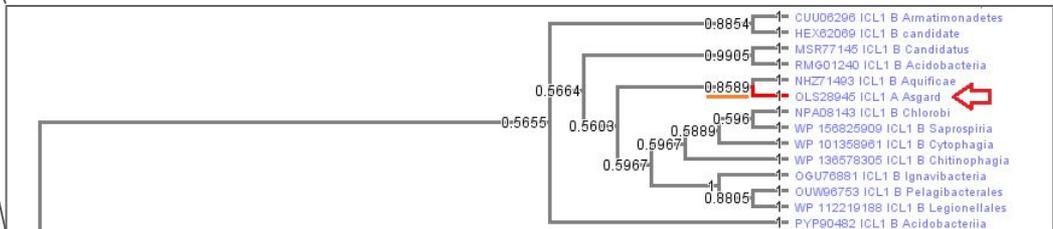
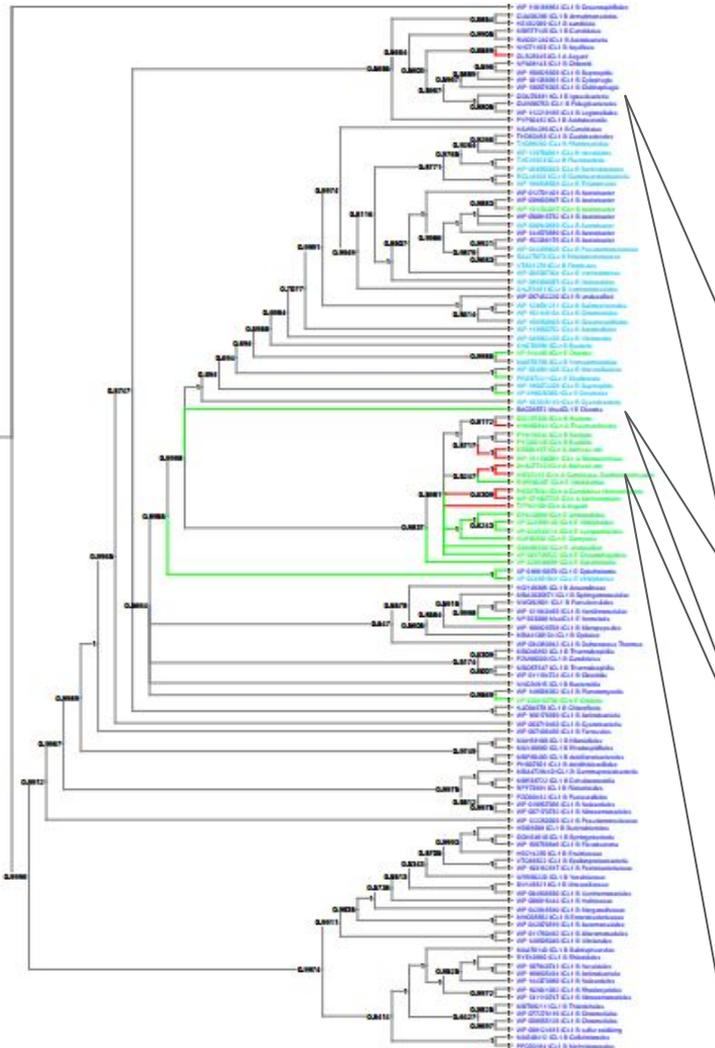
But checking pairwise alignments confirms that it is contamination - the sequences are just too similar. Thus, no horizontal gene transfer was found.

	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		HVRAVKEKLGST--WGAIKRPEDAARMVQNRFHGLDIAKYTAAIMRKMAYEDADSSKY H++A+ E G W AI PE AARM +QNR +GLDIA+YTAAMR+DMAEYD D+S+Y			66
Sbjct	7		HIKALTELAGKNPNWAAINPEYAARMRLQNRKFSGLDIARYTAAIMRRDMAEYDQASRY			66
Query	67		TQSLGCWHGFI AQQKMIANKYFGTTSKRYIYLSGWMVAALRSEFGPLPDSMHEKTSVP TQSLGCWHGFI AQQK+IA KK+ +T KRY+YLSGWM+AALRSEFGPLPDSMHEKT+VP			126
Sbjct	67		TQSLGCWHGFI AQQKLI AIKKHQKSTEKRYLYLSGWMIAALRSEFGPLPDSMHEKTAVP			126
Query	127		KLIAEYTFILRQADAKELNDLFRALQKAEQAGDSAKAAEII SQIDNFESHVVPITADIDA LI EY FLRQAD +EL+ LF AL KA AG++ + +AQI+ F+HHVPIIADIDA			186
Sbjct	127		ALIGEYDFLRQADTREL DLLFSALDKARAAGETREETALLAQIEGFQTHVVPITADIDA			186
Query	187		FGFNEEATYLLTKQMI EAGACATQIENQVSDAKQCQGHQGVKTVPHEDFLAKINAVRYAF FGFN EATYLL K+MIEAGAC IQIENQVSD KQCQGH QKVTVPH DFLAKINAVRYAF			246
Sbjct	187		FGFNAEATYLLAKKMI EAGACCQIENQVSD EKCQGHQGVKTVPHADFLAKINAVRYAF			246
Query	247		LELGDVDEGVIVARTDSEADLTQKIPVSKKEDGLASQYISYLDTKEIDISEASDEILIK LELGDV+GVIVARTDS GA LT++I VS++ GDL QY +LD EI +E +E+LI			306
Sbjct	247		LELGDVDDGVIVARTD SLGAGLTKQYAVSEKPGDLGDQYNRFLDCDEIHPAEIQNGEVLTIS			306
Query	307		RDGKLRHPTRLASGLYQFREGTQHDRVLDVCTS LQNGADMIWIETPTPDVAGIAGFVND RDGKL RP RLASGL+QF++G+ DR VLDCT+SLQNGADMIWIET P + I G ++			366
Sbjct	307		RDGKLRPKRLASGLFQFQKSGSEDRCLVDCITSLQNGADMIWIETETKPHIGQIKGMLDR			366
Query	367		IKKQVPAKLVYNNPSPFNWTLNFRQQAYDRWVAEGKDVSGVDRAKLMSAEYDNLSELAAD I++ VPAKLVYNNPSPFNWTLNFRQQ +D W AEG+D++ Y R LMSAEYD +EL			426
Sbjct	367		TREVVPAKLVYNNPSPFNWTLNFRQQVDAQEGRDLAAYQRDALMSAEYDTELKGVK			426
Query	427		ADEKIRTFQADAAREAGVFHHLITLPTTYHTAALSTHELAKYGFGEGLMAYYAGVQRKEI AD KI++FQ DA+R+AGVFHHLITLPTTYHTAALST L++GYFG +MLAYV GVQR+EI			486
Sbjct	427		ADAKIQSFQRDASRDAGVFHHLITLPTTYHTAALSTDLMSRGVFGSDAMLAYYKGVQRQREI			486
Query	487		RGGIACVKHQAMAGSDIGDDHKEIFAGENALKAGDSKNTMNOF 530 R GIACVKHQ MAGSDIGD+HKE FAG ALKA NTH+QF			529
Sbjct	487		RRGIACVKHQMAGSDIGDHNHKEYFAGVAALKAS -GKDNTHMQF			529

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

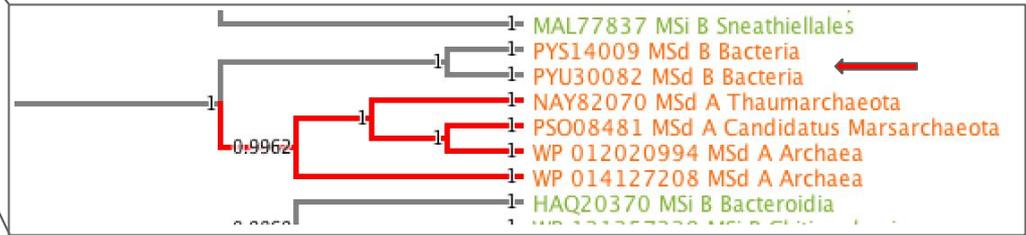
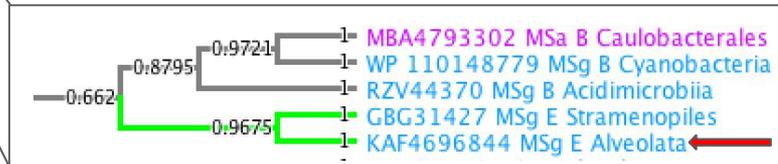
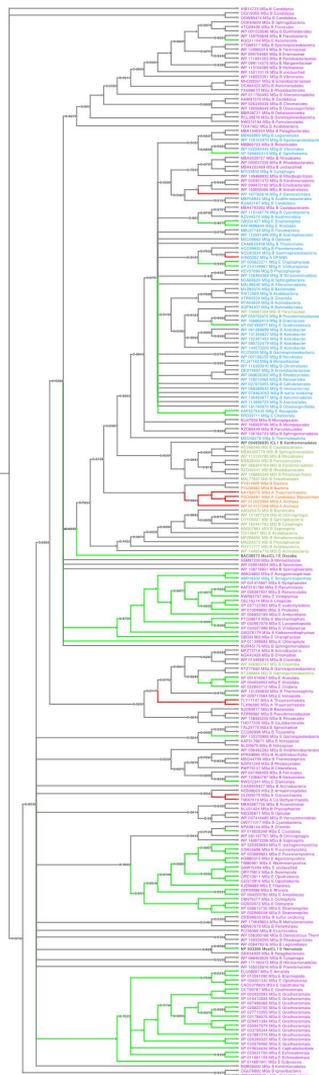
# ICL

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.