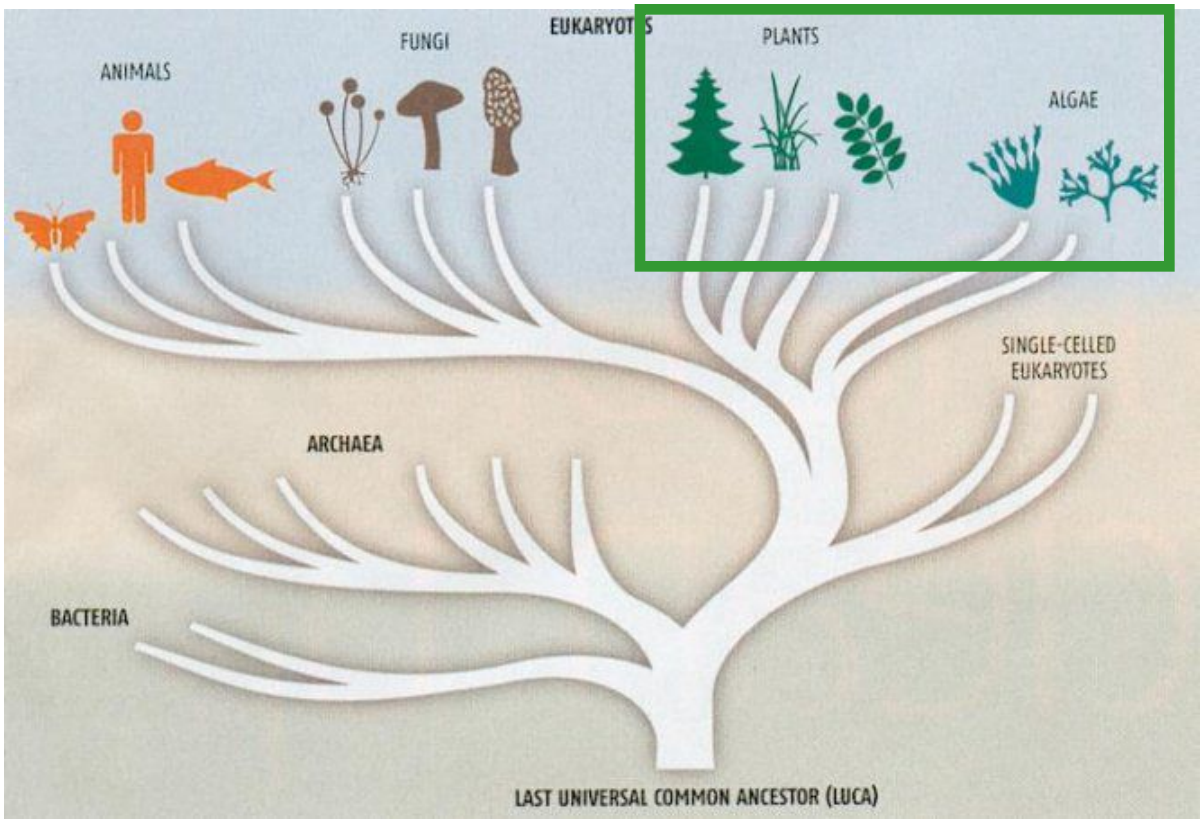
ГЕНОМ ПРО- И ЭУКАРИОТ

Геном эукариот

Новосибирск 2008

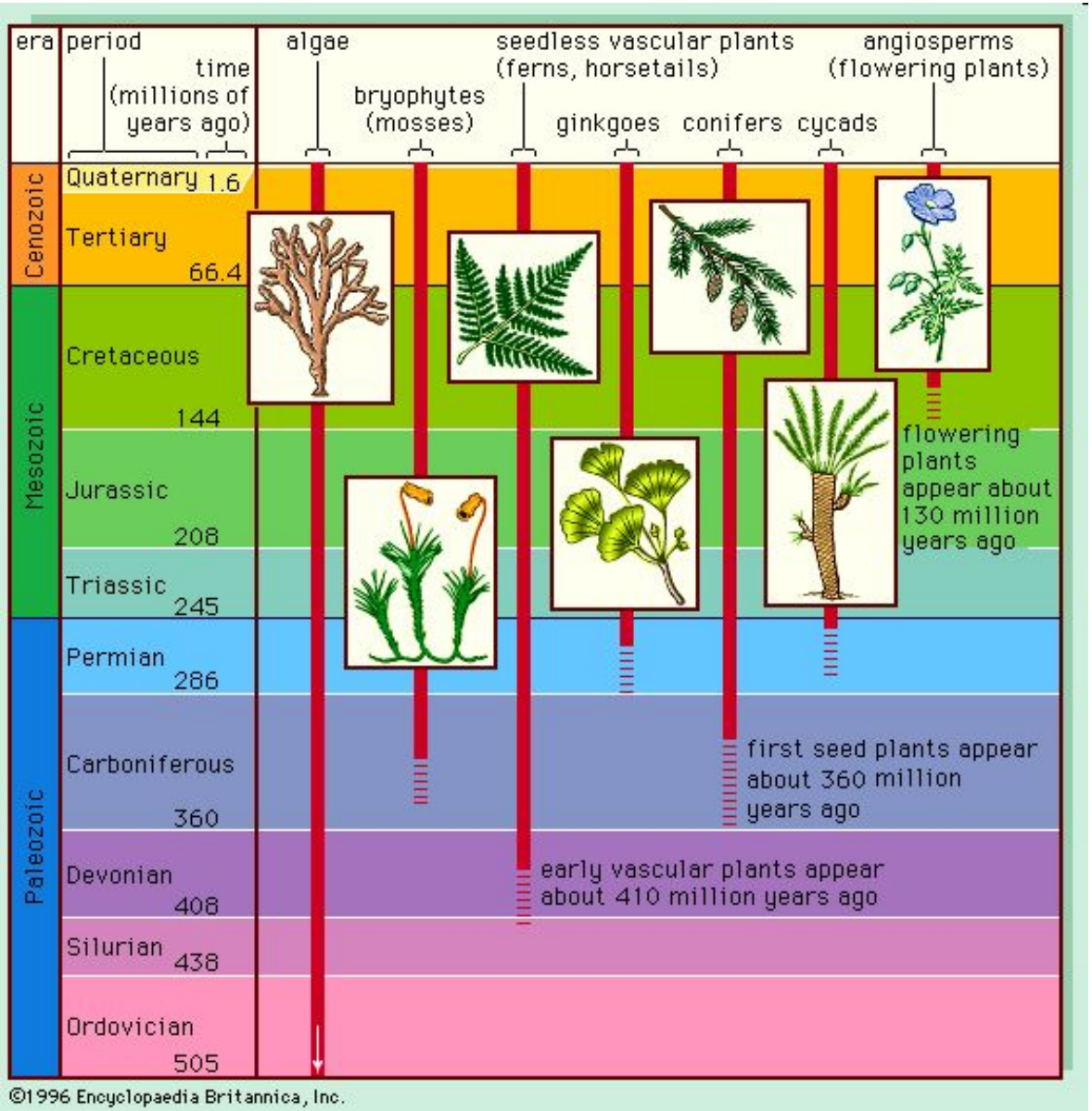
Древо жизни - растения



Растения (лат. Plantae или лат. Vegetabilia) - одна из основных групп многоклеточных организмов. По данным 2004 года было зарегистрировано около 287 655 различных видов растений, среди них около 258 650 цветковых, 16 000 мхов, 11 000 папоротниковиков и 8000 зелёных водорослей.

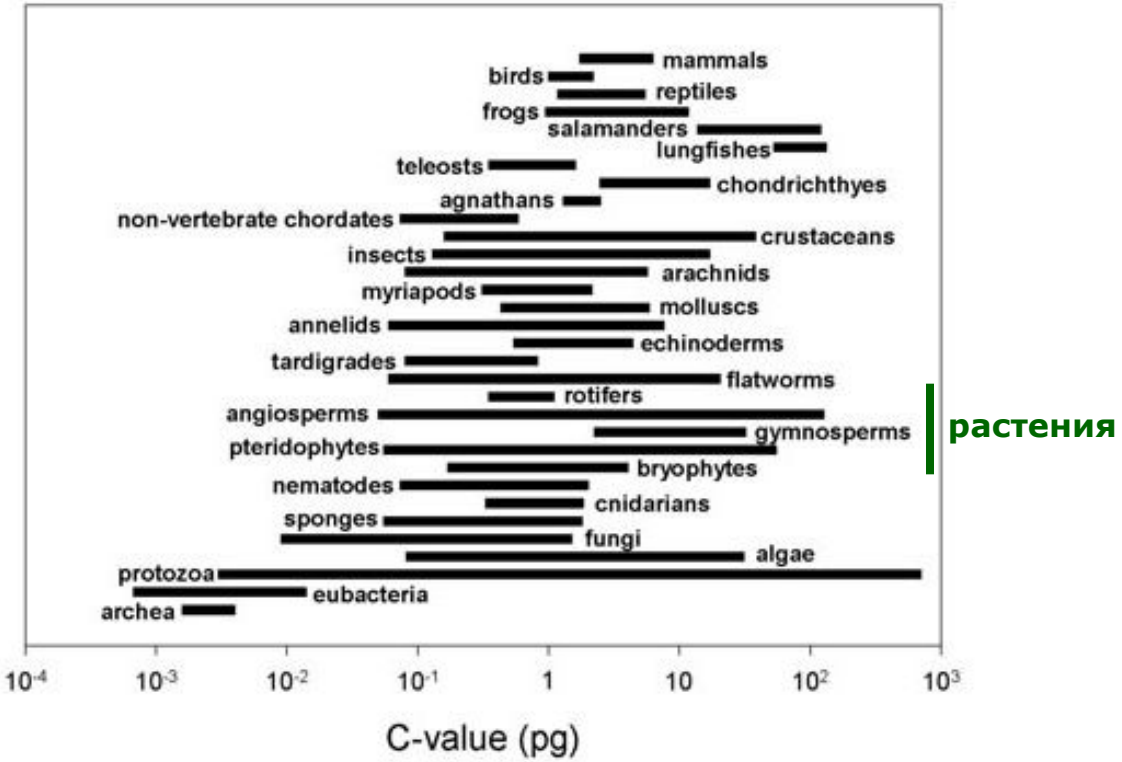
В настоящее время определяющими признаками растений являются наличие плотной, не пропускающей твёрдых частиц, клеточной оболочки (как правило, состоящей из целлюлозы), а также автотрофное (обычно фототрофное) питание — получение органических веществ с помощью углекислого газа и энергии солнца (фотосинтез). Однако есть растения-паразиты, неспособные к фотосинтезу.

Древо жизни - растения



Эволюция растений

Растения – геном



Размеры геномов растений

Растения – геном

Common name	Scientific name	Nuclear genome size (1)
Wheat	*Triticum aestivum*	15,966
Onion	*Allium cepa*	15,290
Garden pea	*Pisum sativum*	3,947
Corn	*Zea mays*	2,292
Asparragus	*Asparagus officinalis*	1,308
Tomato	*Lycopersicum esculentum*	907
Sugarbeet	*Beta vulgaris*	758
Apple	*Malus X domestica*	743
Common bean	*Phaseolus vulgaris*	637
Cantaloupe	*Cucumis melo*	454
Grape	*Vitis vinifera*	483
Man	*Homo sapiens*	2,910

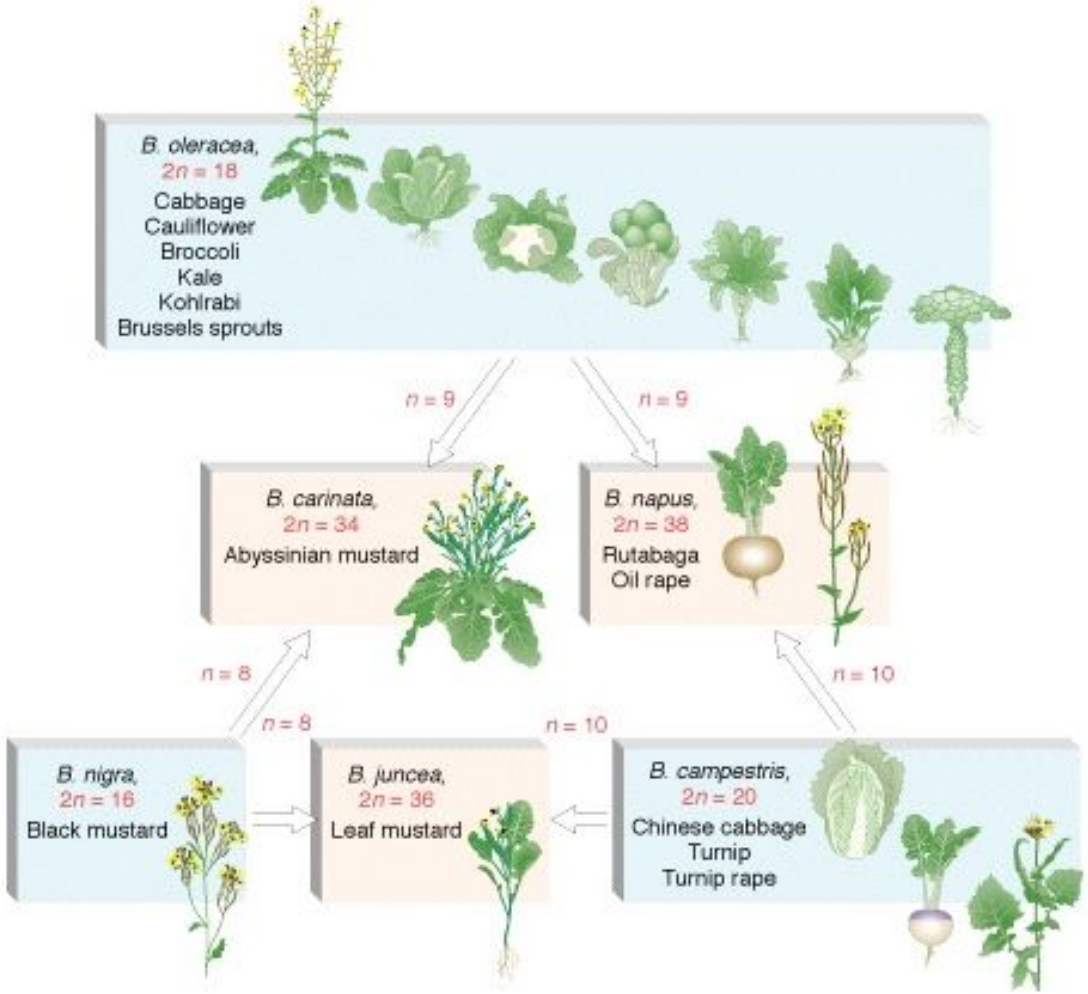
1: Expressed in Megabases (1Mb:1,000,000 bases)

Размеры геномов растений могут отличаться в сотни раз

Растения – размер генома

Самым коротким геномом среди растений (и эукариот) обладает зеленая водоросль *Ostreococcus tauri*. Размер генома этого планктонного микроорганизма **всего около 11.500.000 п.н.** В геноме было обнаружено ~8.000 генов, около 60% не имеют сходства с известными генами других организмов, 20% генов гомологичны генам других зеленых водорослей, 20% - бактериальным генам. Судя по всему, горизонтальный перенос генов играет существенную роль в эволюции геномов зеленых водорослей.

Растения – размер генома



Роль полиплоидизации в эволюции некоторых культурных растений

Полиплоидия, в частности аллополиплоидия, играет существенную роль в эволюции растений. Аллополиплоидия - кратное увеличение числа хромосом у гибридов при скрещивании разных видов (межвидовая и межродовая гибридизация). У аллополиплоидов размер генома может быть огромен – до **20 миллиардов п.н.**

Растения – геном

NCBI ENTREZ Genome Project connection information discovery

Entrez | Genome | Gene | Nucleotide | Protein | PopSet | PubMed | Taxonomy

Search Genome Project Go Clear

Properties of Eukaryotic Genome Sequencing Projects

Organism Group: -- All Plants -- Sequencing Status: All Sequencing Method: All Go Reset

Abbreviations: GB - GenBank Accessions; PM - PubMed; R - RefSeq Accessions; G - Entrez Gene; T - Trace Archive; B - BLAST; M - Map Viewer; F - FTP Sites

48 Eukaryotic Genome Sequencing Projects Selected: Complete - 3, Assembly - 10, In Progress - 35 save

Organism Information							Sequence Information							Links					
GPID	Organism	Group	Subgroup	TaxID	Genome Size (Mb)	# Chr	Status	Method	Depth	Release Date	Center/Consortium	GB	PM	R	G	T	B	M	F
18745	Aquilegia formosa	Plants	Land Plants	223430			In Progress	WGS			DOE Joint Genome Institute								
15663	Arabidopsis lyrata	Plants	Land Plants	59689			In Progress				DOE Joint Genome Institute						B		
13190	Arabidopsis thaliana	Plants	Land Plants	3702	119.2	5	Complete	WGS & Clone-based		12/14/2000	Arabidopsis Genome Initiative [more]	GB	PM	R	G	T	B	M	F
12480	Brassica oleracea	Plants	Land Plants	3712	600	9	In Progress	WGS	0.44X		TIGR		PM			T			
28961	Brassica rapa Chiifu	Plants	Land Plants	3711	500	10	In Progress	Clone-based			The Multinational Brassica rapa Sequencing Project [more]								
13878	Capsella rubella	Plants	Land Plants	81985	250		In Progress				DOE Joint Genome Institute								
20267	Carica papaya SunUp	Plants	Land Plants	3649			Assembly	WGS	3X		The Papaya Genome Sequencing Consortium [more]				G				
12260	Chlamydomonas reinhardtii	Plants	Green Algae	3055	100	17	Assembly	WGS	12.8X	02/14/2004	DOE Joint Genome Institute	GB	PM		G	T	B		F
18711	Chlorella sp. NC64A	Plants	Green Algae	310507			In Progress	WGS			DOE Joint Genome Institute								

По количеству геномных проектов растения значительно уступают другим группам эукариот. Это связано, в первую очередь, с огромным размером геномов большинства культурных растений, а также с существенными трудностями, которые встречаются при сборке генома – наличие огромного числа повторенных последовательностей (до 90%) делает процесс сборки генома очень сложным.

Растения – геном

NCBI ENTREZ Genome Project

connection information discovery

Entrez | Genome | Gene | Nucleotide | Protein | PopSet | PubMed | Taxonomy

Search Genome Project Go Clear

Properties of Eukaryotic Genome Sequencing Projects

Organism Group: Land Plants Sequencing Status: All Sequencing Method: All Go Reset

Abbreviations: GB - GenBank Accessions; PM - PubMed; R - RefSeq Accessions; G - Entrez Gene; T - Trace Archive; B - BLAST; M - Map Viewer; F - FTP Sites

39 Eukaryotic Genome Sequencing Projects Selected: Complete - 3, Assembly - 8, In Progress - 28 save

	Organism Information						Sequence Information							Links					
GPID	Organism	Group	Subgroup	TaxID	Genome Size (Mb)	# Chr	Status	Method	Depth	Release Date	Center/Consortium	GB	PM	R	G	T	B	M	F
18745	Aquilegia formosa	Plants	Land Plants	223430			In Progress	WGS			DOE Joint Genome Institute								
15663	Arabidopsis lyrata	Plants	Land Plants	59689			In Progress				DOE Joint Genome Institute						B		
13190	Arabidopsis thaliana	Plants	Land Plants	3702	119.2	5	Complete	WGS & Clone-based		12/14/2000	Arabidopsis Genome Initiative [more]	GB	PM	R	G	T	B	M	F

На сегодняшний день в базе NCBI индексируется 39 геномных проектов растений,
3 из них завершены – это геномы модельного организма *Arabidopsis thaliana* и двух сортов риса *Oryza sativa*.

Растения – геном

NCBI ENTREZ Genome Project connection information discovery

Entrez | Genome | Gene | Nucleotide | Protein | PopSet | PubMed | Taxonomy

Search Genome Project Go Clear

Properties of Eukaryotic Genome Sequencing Projects

Organism Group: Green Algae Sequencing Status: All Sequencing Method: All Go Reset

Abbreviations: GB - GenBank Accessions; PM - PubMed; R - RefSeq Accessions; G - Entrez Gene; T - Trace Archive; B - BLAST; M - Map Viewer; F - FTP Sites

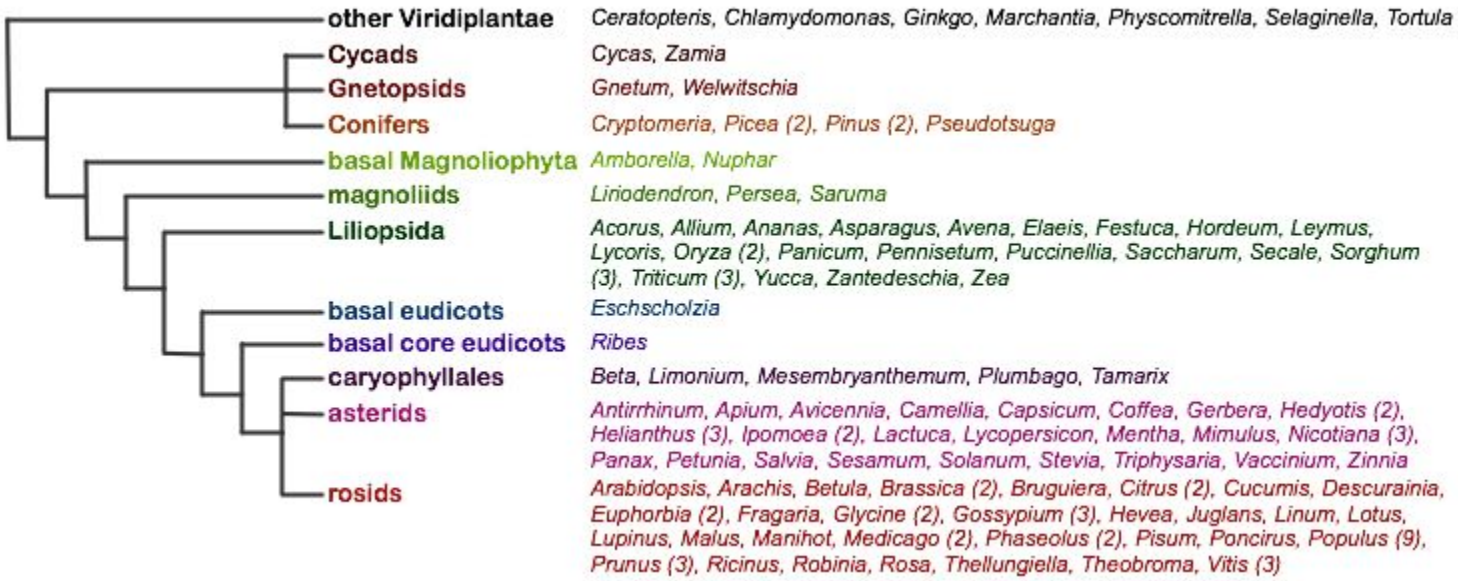
9 Eukaryotic Genome Sequencing Projects Selected: Complete - 0, Assembly - 2, In Progress - 7 save

	Organism Information						Sequence Information							Links					
GPID	Organism	Group	Subgroup	TaxID	Genome Size (Mb)	# Chr	Status	Method	Depth	Release Date	Center/Consortium	GB	PM	R	G	T	B	M	F
12260	Chlamydomonas reinhardtii	Plants	Green Algae	3055	100	17	Assembly	WGS	12.8X	02/14/2004	DOE Joint Genome Institute	GB	PM		G	T	B		F
18711	Chlorella sp. NC64A	Plants	Green Algae	310507			In Progress	WGS			DOE Joint Genome Institute								
18715	Chlorella vulgaris	Plants	Green Algae	3077			In Progress	WGS			DOE Joint Genome Institute				G				F
18609	Dunaliella salina	Plants	Green Algae	3046			In Progress				DOE Joint Genome Institute								
15676	Micromonas pusilla	Plants	Green Algae	38833			In Progress				DOE Joint Genome Institute								
15678	Micromonas pusilla	Plants	Green Algae	38833	15		In Progress				DOE Joint Genome Institute								
13044	Ostreococcus lucimarinus CCE9901	Plants	Green Algae	436017	13.25	21	Assembly	WGS		04/10/2007	DOE Joint Genome Institute [more]	GB			G		B		
12912	Ostreococcus tauri OTH95	Plants	Green Algae	70448	12.5	20	In Progress	WGS & Clone-based	7X	04/30/2005	Laboratoire Arago, France	GB	PM		G				
13109	Volvox carteri	Plants	Green Algae	3067	120		In Progress				DOE Joint Genome Institute					T			

Зеленые водоросли обладают гораздо меньшими геномами, чем другие растения – неудивительно, что из 48 геномных проектов растений – 9 относятся к зеленым водорослям.

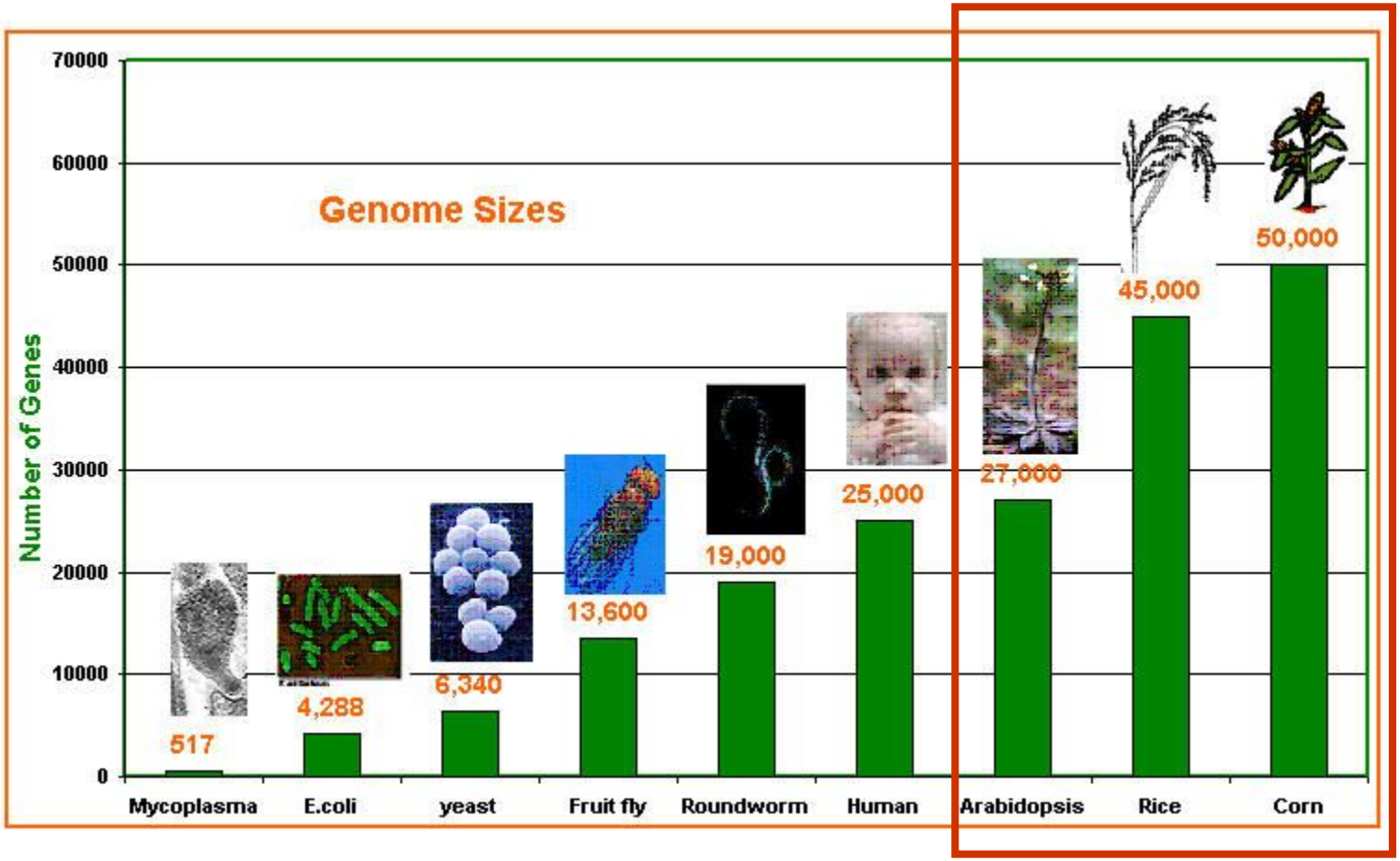
Растения – геном

Вероятно, количество геномных проектов сильно занижено в NCBI.



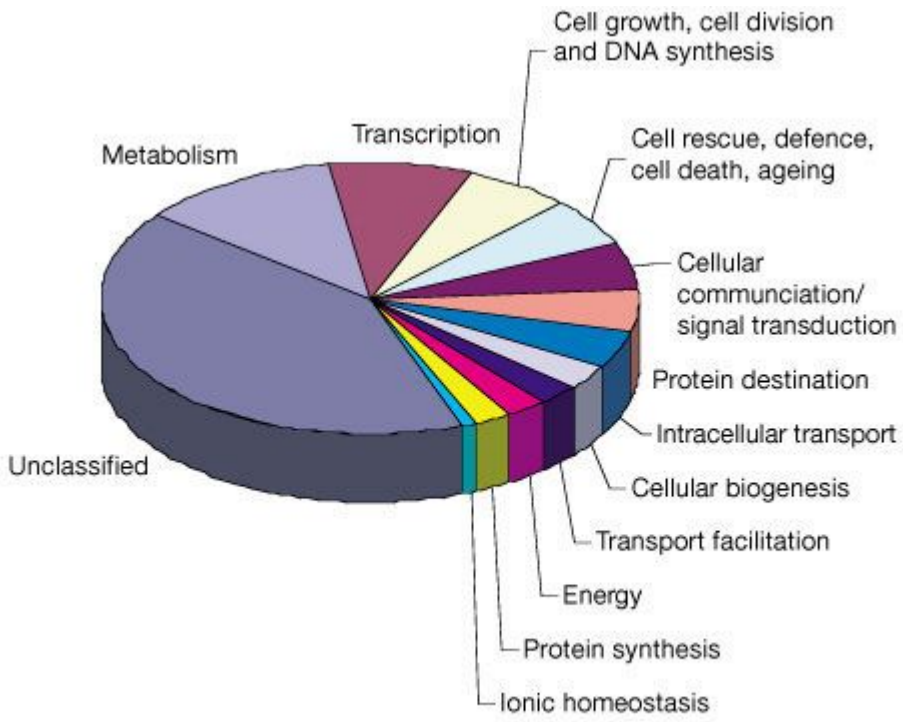
Геномные проекты растений согласно другой базе данных - AgroGenome

Растения – состав генома



Количество генов

Растения – состав генома



Генный состав *Arabidopsis thaliana*

Растения – состав генома

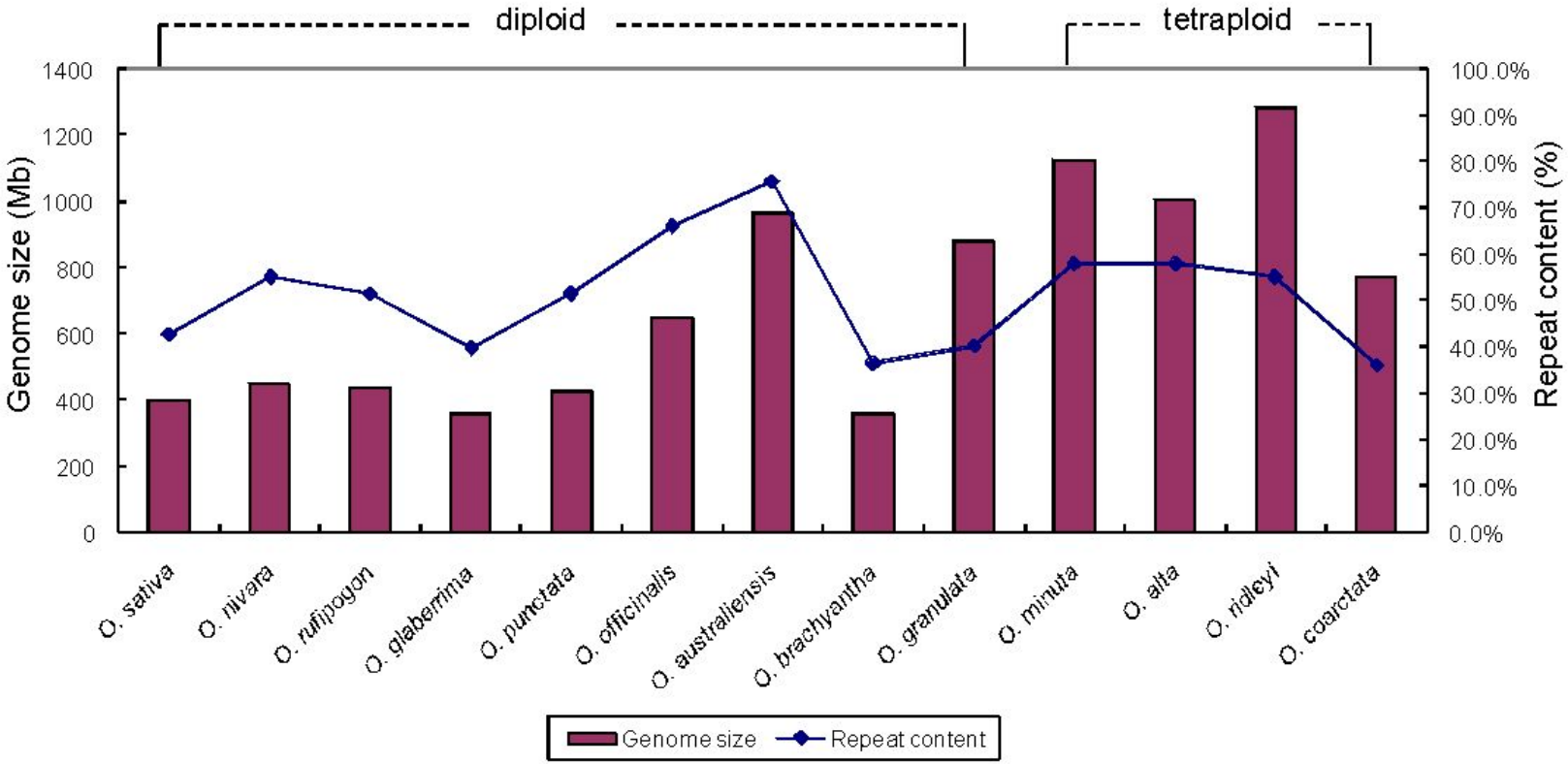
Repeat content analysis of 12 *Oryza* species using RepeatMasker and RECON

Species	Repeat content by RepeatMasker	Repeat content by RECON			Total repeat content
		Total	Overlap with RepeatMasker (% of total)	Uniqe (% of total)	
O. sativa (Nipponbare)	29.4%	37.2%	23.8% (64%)	13.4% (36%)	42.8%
O. nivara	37.0%	47.6%	29.5% (62%)	18.1% (38%)	55.1%
O. rufipogon	36.6%	49.7%	34.6% (70%)	15.2% (30%)	51.7%
O. glaberrima	28.5%	30.9%	19.8% (64%)	11.2% (36%)	39.7%
O. punctata	40.8%	41.4%	30.7% (74%)	10.7% (26%)	51.5%
O. officinalis	47.2%	56.5%	37.6% (67%)	18.9% (33%)	66.1%
O. minuta	44.0%	46.3%	32.2% (70%)	14.1% (30%)	58.1%
O. alta	43.2%	47.9%	33.2% (69%)	14.7% (31%)	57.9%
O. australiensis	55.0%	66.4%	45.4% (68%)	21.0% (32%)	76.0%
O. brachyantha	20.9%	27.5%	11.6% (42%)	15.9% (58%)	36.8%
O. granulata	30.1%	30.9%	20.5% (66%)	10.4% (34%)	40.5%
O. ridleyi	35.6%	42.7%	23.0% (54%)	19.7% (46%)	55.2%
O. coarctata	19.7%	24.5%	7.7% (32%)	16.7% (68%)	36.4%

Kim *et al. Genome Biology* 2008 **9**:R45 doi:10.1186/gb-2008-9-2-r45

Анализ повторов 13 видов риса.
Значительную часть геномов растений занимают повторенные последовательности.

Растения – состав генома



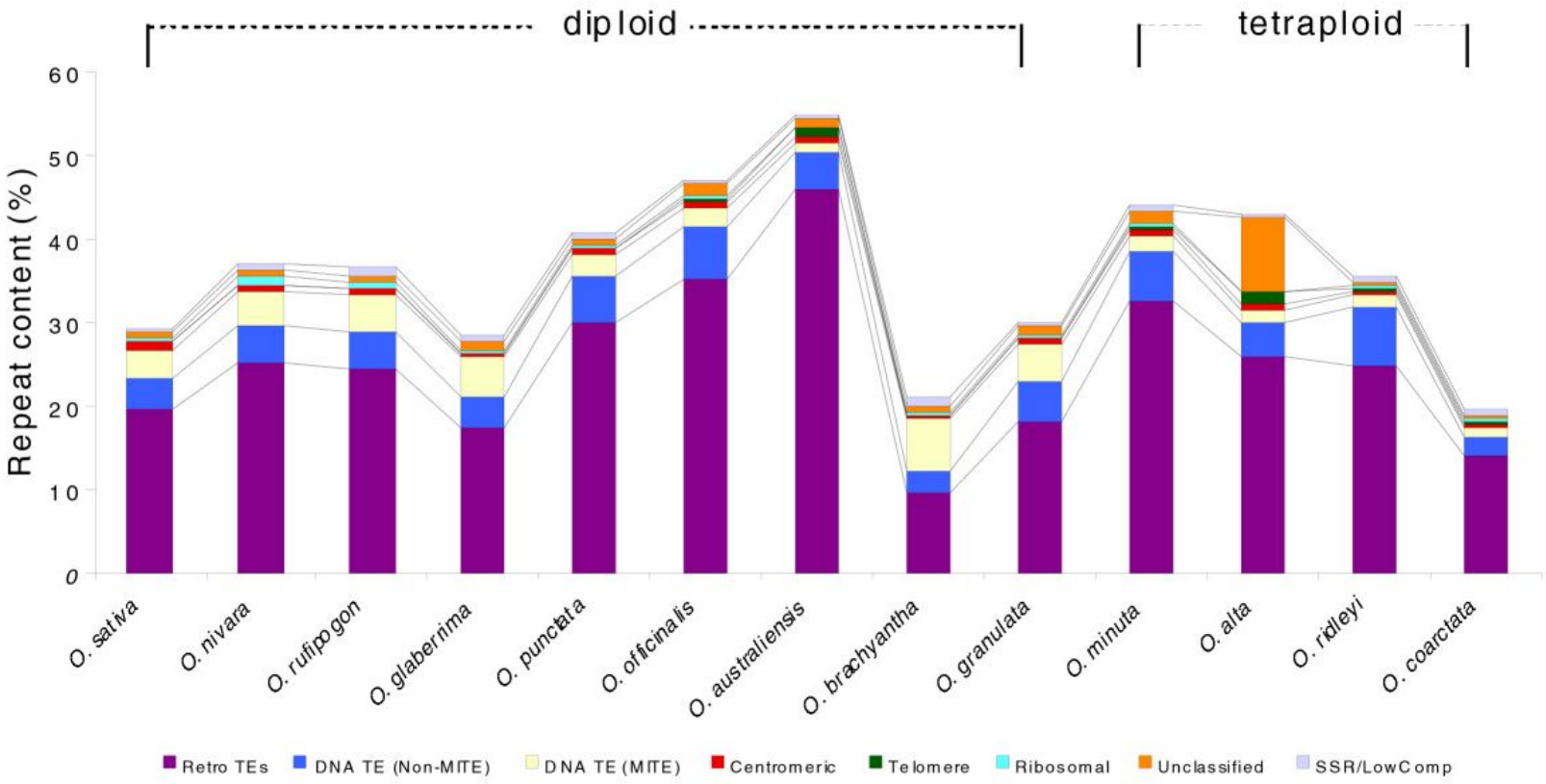
Корреляция размеров генома 13 видов риса и количества повторов

Растения – мобильные элементы

Class/subclass/ superfamily	Species	Autonomous member(s)	Non-autonomous member(s)	Copy number of the entire family
Class 1				
Non-LTR retrotransposons				
LINEs; *L1* clade	*Z. mays*	*Cin4*	–	50–100
	L. speciosum	*Del2*‡	–	250,000
	A. thaliana	*Tal1*‡	–	1–6
SINEs	*N. tabacum*	–	*TS*	50,000
	B. napus	–	*S1*	500
LTR retrotransposons				
copia-like	*N. tabacum*	*Tnt1A*	–	>100
	N. tabacum	*Tto1*	–	30 (300)
	Hordeum sp.	BARE-1	–	5,000–22,000
	O. sativa	*Tos17*	–	2–5 (30)
	Z. mays	*Hopscotch*	–	5–8
	Z. mays	*Opie-2*‡	–	100,000
	Z. mays	–	*BS1*	1–5
gypsy-like	*Z. mays*	*Magellan*	–	4–8
	Z. mays	*Huck-2*‡	–	200,000
	O. sativa	*RIRE2* ?	*Dasheng*	1,200
	A. thaliana	*Athila 4*‡	–	22
	A. thaliana	*Athila 6*‡	–	11
	A. thaliana	*Ta3*	–	1
	A. thaliana	*Tar17*	–	2
Class 2				
DNA transposons				
hAT	*Z. mays*	*Ac*	*Ds*	50–100
CACTA	*Z. mays*	*Spm*	*dSpm*	50–100
	A. thaliana	*CAC1*	*CAC2*	4 (20)
Mutator	*Z. mays*	*MuDR*	*Mu1*	10–100
	A. thaliana	*AtMu1*	–	1 (4)
PIF/*Harbinger*	*Z. mays*	*PIFa*	*mPIF*	6,000
	Angiosperms	*PIF*-like	*Tourist*-like	Variable
Tc1/*mariner*	Angiosperms	*MLE*s	*Stowaway*-like	Variable

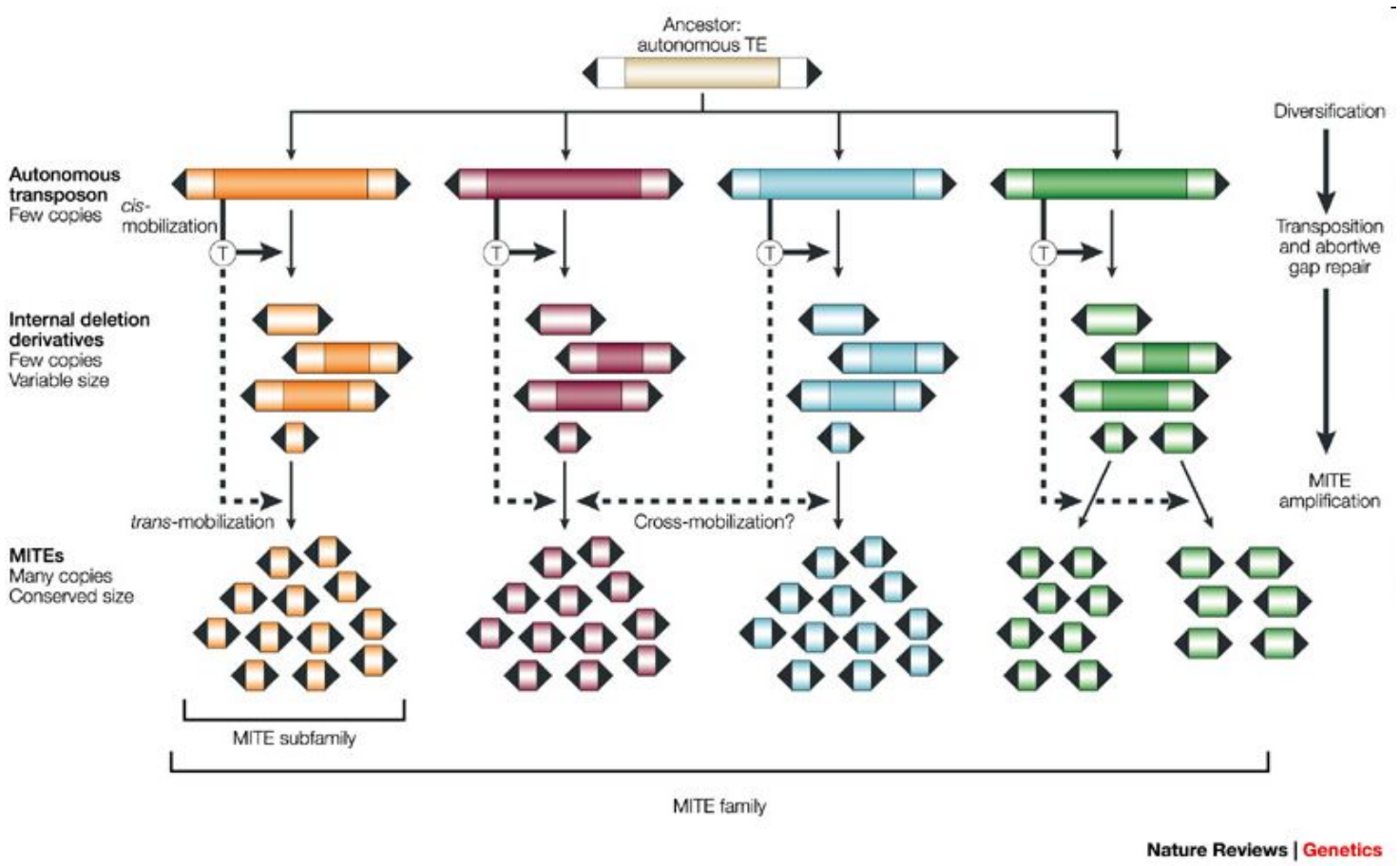
Все классы и типы мобильных элементов представлены в геномах растений. Их копийность может значительно варьировать, однако, как правило значительную часть всех мобильных элементов в геноме составляют LTR ретротранспозоны.

Растения – мобильные элементы



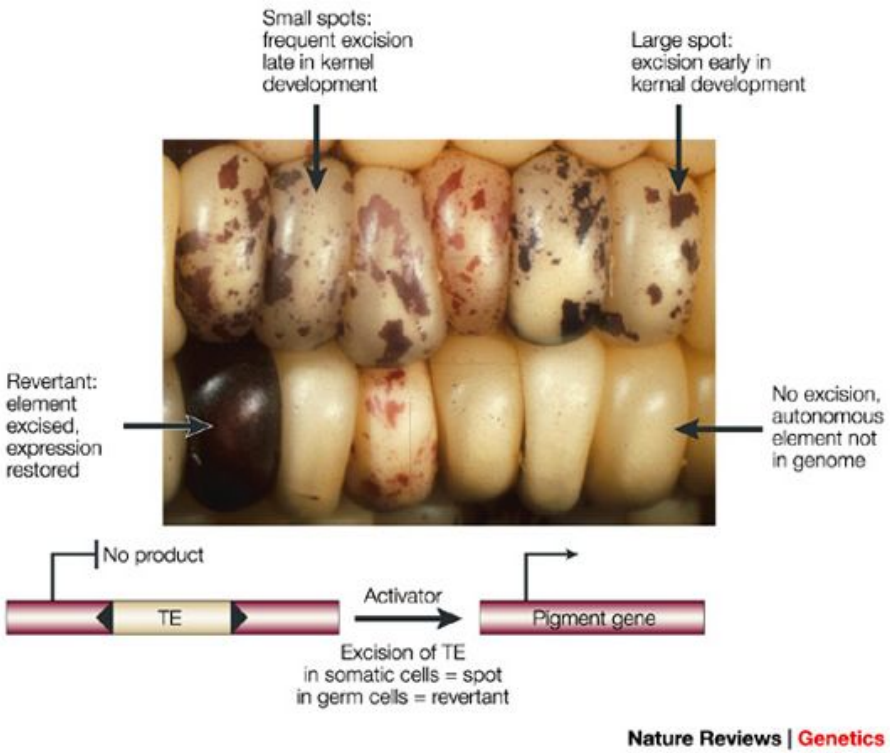
Сравнительный анализ состава повторенной ДНК 13 видов риса.
Наибольшее количество повторов представляют собой ретротранспозоны,
ДНК транспозоны и MITE элементы.

Растения – мобильные элементы



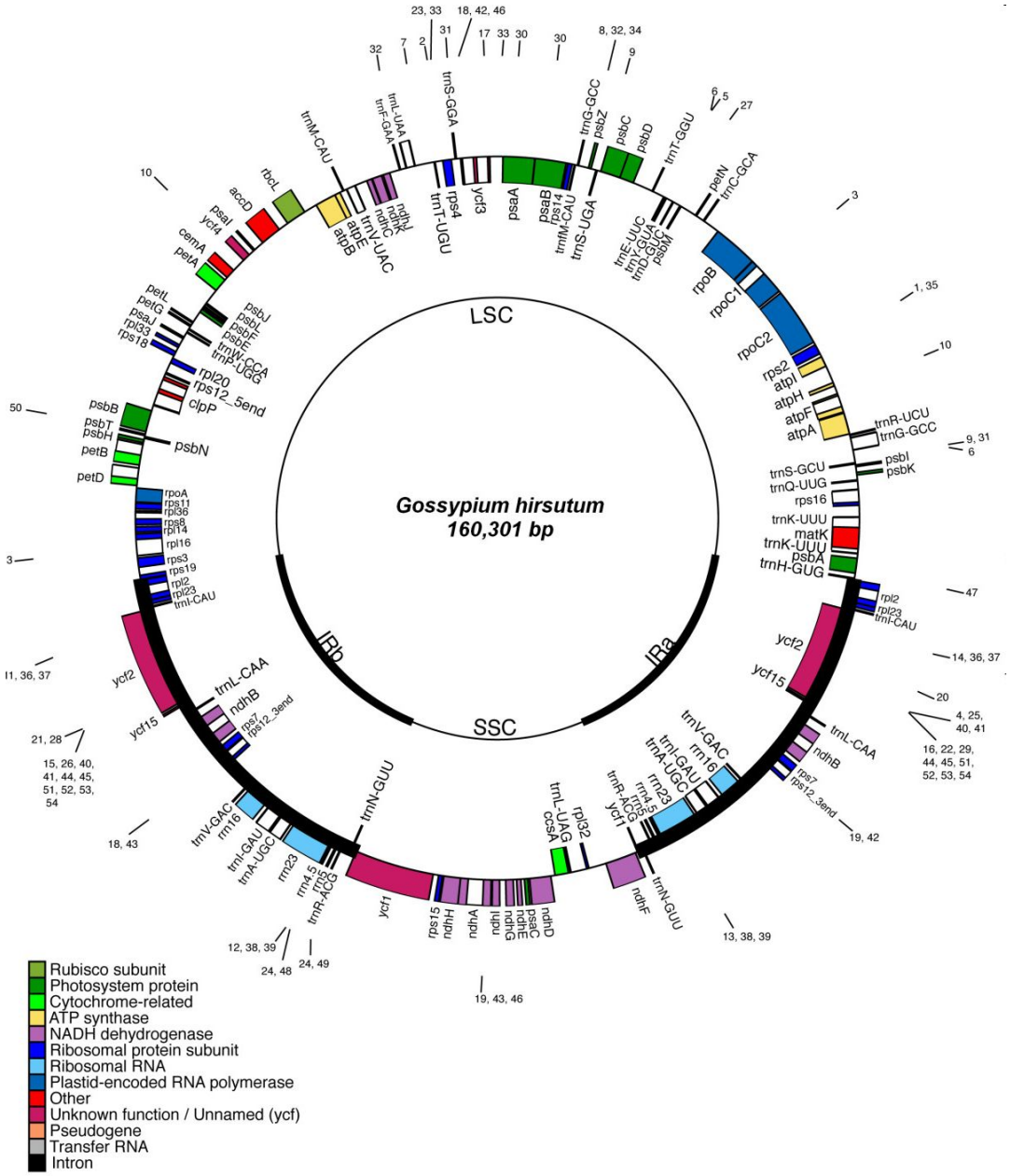
Возникновение MITE элементов.

Растения – мобильные элементы



Мобильные элементы играют существенную роль в эволюции растений.

Растения – хлоропластный геном



Sizes of completely sequenced chloroplast and related plastid genomes.[1]

Taxonomic characterization and species	Size (kb)
Kingdom Protoctista	
Astasia longa	73.3
Euglena gracilis	143.2
Toxoplasma gondii	35.0
Algae	
Chlorella vulgaris	150.6
Cyanidium caldarium	164.9
Cyanophora paradoxa	135.6
Guillardia theta	121.5
Mesostigma viride	118.4
Nephroselmis olivacea	200.8
Porphyra purpurea	191.0
Odontella sinensis	119.7
Bryophyta	
Marchantia polymorpha	121.0
Conifer	
Pinus thumbergii	119.7
Monocots, Poales	
Oryza sativa	134.5
Triticum aestivum	134.5
Zea mays	140.4
Eurosids	
Fabales	
Lotus japonicus	150.5
Myrtales	
Oenothera elata	163.9
Brassicales	
Arabidopsis thaliana	154.5
Euasterids	
Solanales	
Nicotiana tabacum	155.9
Lamiales	
Epifagus virginiana	70.0

[1] Source: Saccone and Pesole (2003).