GoFORSYS

Universität Potsdam

Max Planck Institute of Molecular Plant Physiology

Max Planck Institute of Colloids and Interfaces

GoFORSYS

C5-Sy25-i008-R

Genetic variability of Calvin cycle genes in *Arabidopsis* accessions

Madlen Stange, Sandra Schwarte, Fanny Wegner, Ralph Tiedemann

Evolutionary Biology, University of Potsdam, Potsdam, Germany

email: madlen.stange@uni-potsdam.de

Introduction

The Calvin cycle forms the basis for the generation of carbohydrates in all eukaryotes. Every year, $5x10^{11}$ tons of CO_2 are fixed by this primary pathway. The Calvin cycle plays a central role in plant metabolism by providing intermediates for starch and sucrose metabolism, as well as isoprenoid and shikimic acid metabolism. All envolved enzymes are located within the stroma of chloroplasts, therefore all nuclear encoded Calvin cycle genes exhibit a chloroplast transit peptide (cTP) for the transport from the nucleus to the chloroplast. In higher plants, the whole cycle consists of 11 enzymes, of which 2 enzymes are bi-specific (SFBA, TKL). The involved enzymes are Rubisco, PGK, GAPDH, TPI, SFBA, FBPase, SBPase, TKL, RPE, RPI, and PRK. Altogether these enzymes catalyze 13 reactions. We sequenced 14 genes via classical Sanger sequencing coding for the above mentioned 11 enzymes and two additional genes coding for regulatory enzymes, Rubisco activase (RCA) and Rubisco methyl transferase (RMT), in 26 worldwide distributed *A. thaliana* accessions, colonizing manifold habitats. We analyzed genetic variation among these 26 accessions, regarding number of substitutions in structural and functional domains, nucleotide diversity, and haplotype diversity. We found the genes being as variable as secondary metabolism genes with some accessions exhibiting possible protein structure damaging non-synonymous substitutions in functional domains.

Results – Gene specific genetic variation

Table 1.1 Multidomain analysis of respective genes, *AtRPI* (*At3g04790*). Analyzed were promoter sequence, entire gene, sum of all exons (number in brackets) and all introns, the chloropastidal transit peptide and individual functional domains predicted by KEGG.

domain	sites	*S*	syn	nonsyn	π
promoter	380	4	-	-	0.0023
gene/exon (1)	834	14	11	3	0.0027
cTP	111	0	0	0	0.0000
rpi A	540	9	8	1	0.0027

Selection — **Z test:** purifying (3.105**); **Tajima's D:** -1.302

Gene tree — random substitutions, no clusters

(** significant at 0.01%, * significant at 0.05%)

Table 1.2 *AtPRK* (*At1g32060*).

domain	sites	*S*	syn	nonsyn	π
gene	1742	88	40	2	0.0109
all exons (5)	1188	42	40	2	0.0075
all introns	554	46	-	-	0.0184
cTP	162	8	7	1	0.0122
phosphoribulokinase	597	23	23	1	0.0080

Selection — **Z test:** purifying (5.486**); **Tajima's D:** -0.720

Gene tree — random substitutions, no clusters

Table 1.3 *AtRCA* (*At2g39730*).

domain	sites	S	syn	nonsyn	π
promoter	1018	7	-	-	0.0009
gene	2399	13	4	0	0.0008
all exons (6)	1425	4	4	0	0.0002
all introns	974	9	-	-	0.0018
cTP	174	0	0	0	0.0000
AAA ATPase	426	1	1	0	0.0002

Selection — **Z test:** purifying (2.184*); **Tajima's D:** -1.887*

Gene tree — random substitutions, no clusters

Table 1.4 *AtRMT* (*At1g14030*).

domain	sites	*S*	syn	nonsyn	π
promoter	1115	62	-	-	0.0098
gene	1975	56	19	15	0.0030
all exons (6)	1458	34	19	15	0.0027
all introns	517	22	-	-	0.0039
cTP	180	8	2	6	0.0086
SET domain	618	17	14	3	0.0027
rubisco LSMT substrate binding	381	6	3	3	0.0014

Selection — **Z test:** purifying (1.661*); **Tajima's D:** -2.085*

Gene tree — random substitutions, no clusters

Table 1.5 *AtPGK* (*At1g56190*).

domain	sites	*S*	syn	nonsyn	π
gene	2052	17	3	6	0.0001
all exons (6)	1437	9	3	6	0.0007
all introns	615	8	-	-	0.0012
cTP	285	4	2	2	0.0018
phosphoglycerate kinase	1143	4	0	4	0.0003

Selection — **Z test:** neutral; **Tajima's D:** -1.910*

Gene tree — random substitutions, no clusters

Table 1.6 *AtPGK* (*At3g12780*).

domain	sites	*S*	syn	nonsyn	π
promoter	880	22	-	-	0.0045
gene	2041	34	10	4	0.0023
all exons (6)	1446	15	10	4	0.0015
all introns	595	19	-	-	0.0043
cTP	225	10	5	4	0.0059
phosphoglycerate kinase	1149	5	5	0	0.0008

Selection — **Z test:** purifying (2.613**); **Tajima's D:** -1.518

Gene tree — 1 major haplogroup

Table 1.7 *AtGAPDH-a1* (*At3g26650*).

domain	sites	*S*	syn	nonsyn	π
promoter	983	52	-	-	0.0251
gene	1635	14	4	0	0.0027
all exons (5)	1191	4	4	0	0.0012
all introns	444	10	-	-	0.0068
cTP	135	1	1	0	0.0011
NAD-binding domain	454	2	2	0	0.0018
active site	24	1	1	0	0.0171
C-terminal domain	471	0	0	0	0.0000

Selection — **Z test:** purifying (1.937*); **Tajima's D:** 0.834

Gene tree — 2 distinct haplogroups

Table 1.8 *AtGAPDH-a2* (*At1g12900*), 2nd splice variant.

domain	sites	*S*	syn	nonsyn	π
promoter	884	11	-	-	0.0020
gene	1114	6	2	2	0.0005
all exons (3)	954	4	2	2	0.0004
all introns	160	2	-	-	0.0010
cTP	144	1	0	1	0.0005
NAD-binding domain	198	0	0	0	0.0000
active site	24	0	0	0	0.0000
C-terminal domain	471	3	2	1	0.0006

Selection — **Z test:** neutral; **Tajima's D:** -1.707

Gene tree — random substitutions, no clusters

Table 1.9 *AtGAPDH-b* (*At1g42970*).

domain	sites	*S*	syn	nonsyn	π
promoter	475	12	-	-	0.0028
gene	2036	61	26	5	0.0003
all exons (9)	1344	31	26	5	0.0029
all introns	692	30	-	-	0.0043
cTP	135	5	0	0	0.0029
NAD-binding domain	459	15	13	2	0.0034
active site	24	1	1	0	0.0113
C-terminal domain	474	7	7	0	0.0035
CP12 domain	63	0	0	0	0.0000

Selection — **Z test:** purifying (4.102**); **Tajima's D:** -1.936*

Gene tree — random substitutions, no clusters

Table 1.10 *AtTPI* (*At2g21170*), 2nd splice variant.

domain	sites	*S*	syn	nonsyn	π
promoter	1177	22	-	-	0.0065
gene	2061	27	4	5	0.0030
all exons (8)	921	9	4	5	0.0017
all introns	1140	18	-	-	0.0040
cTP	174	4	2	2	0.0050
triosephosphate isomerase	693	5	2	3	0.0011
active site	33	0	0	0	0.0000

Selection — **Z test:** neutral; **Tajima's D:** -0.505

Gene tree — random substitutions, no clusters

Table 1.11 *AtSFBA* (*At2g21330*), 2nd splice variant.

domain	sites	S	syn	nonsyn	π
promoter	1020	31	-	-	0.0051
gene	1397	33	9	1	0.0043
all exons (5)	936	10	9	1	0.0016
all introns	461	23	-	-	0.0103
cTP	30	0	0	0	0.0000
fructose bisphosphate aldolase class I (FBAI)	762	7	6	1	0.0013
FBAI active site	33	0	0	0	0.0000

Selection — **Z test:** purifying (2.829**); **Tajima's D:** -1.648

Gene tree — random substitutions, no clusters

Table 1.12 *AtSFBA* (*At4g38970*), 2nd splice variant.

domain	sites	S	syn	nonsyn	π
promoter	1065	19	-	-	0.0042
gene	1911	44	4	1	0.0051
all exons (7)	1146	5	4	1	0.0005
all introns	765	39	-	-	0.0123
cTP	138	0	0	0	0.0000
fructose-bisphosphate aldolase class I (FBAI)	1038	4	3	1	0.0005
FBAI active site	33	0	0	0	0.0000

Selection — **Z test:** neutral; **Tajima's D:** -1.583

Gene tree — 2 distinct haplogroups (1 major, 1 minor)

Table 1.13 *AtFBPase* (*At3g54050*).

domain	sites	*S*	syn	nonsyn	π
promoter	1072	26	-	-	0.0059
gene	1578	29	12	2	0.0081
all exons (4)	1254	14	12	2	0.0052
all introns	324	15	-	-	0.0190
cTP	171	4	3	1	0.0118
FBPase	1002	9	9	0	0.0041
active site	39	2	2	0	0.0190

Selection — **Z test:** purifying (3.475**); **Tajima's D:** 2.717**

Gene tree — 2 distinct haplogroups

Table 1.14 *AtSBPase* (*At3g55800*).

domain	sites	*S*	syn	nonsyn	π
promoter	371	14	-	-	0.0120
gene	1791	61	21	5	0.0098
all exons (8)	1182	26	21	5	0.0064
all introns	609	35	-	-	0.0167
cTP	177	5	2	2	0.0077
FBPase	930	16	14	2	0.0050
active site	39	3	3	0	0.0249

Selection — **Z test:** purifying (4.440**); **Tajima's D:** 0.429

Gene tree — 2 distinct haplogroups (1 major, 1 minor)

Table 1.15 *AtTKL* (*At3g60750*).

domain	sites	*S*	syn	nonsyn	π
promoter	690	3	-	-	0.0012
gene	2821	14	6	0	0.0043
all exons (7)	2226	6	6	0	0.0005
all introns	595	8	-	-	0.0018
cTP	195	0	0	0	0.0000
PP	993	0	0	0	0.0000
Pyr	516	4	4	0	0.0010
C-term	321	2	2	0	0.0019

Selection — **Z test:** purifying (1.913*); **Tajima's D:** -0.869

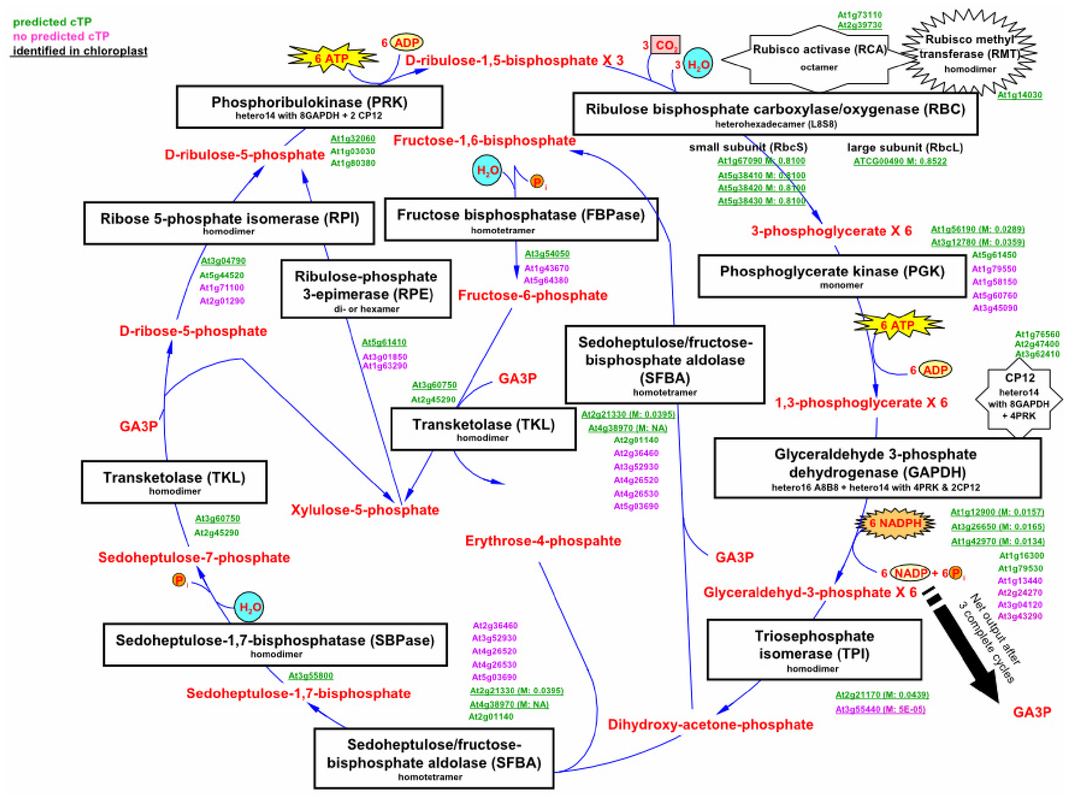
Gene tree — random substitutions, no clusters

Table 1.16 *AtRPE* (*At5g61410*).

domain	sites	*S*	syn	nonsyn	π
promoter	1021	41	-	-	0.0071
gene	1753	18	7	0	0.0038
all exons (9)	846	7	7	0	0.0028
all introns	907	11	-	-	0.0048
RPE family	603	4	4	0	0.0023

Selection — **Z test:** purifying (2.487**); **Tajima's D:** 1.422

Gene tree — 2 distinct haplogroups



Peltier J et al. Mol Cell Proteomics 2006; 5:114-133, modified

Results – Implications on functional domains

Table 17. Selected non-synonymous substitutions and their impact on protein structure predicted by PolyPhen2 Server.

gene	nonsynonymous substitution	impact on protein structure	score	position in gene	affected accessions
AtGAPDH-a2 (*At1g12900*)	T272S	possibly damaging	0.620	C-terminal domain	Cvi-0
AtRMT (*At1g14030*)	S24C	possibly damaging	0.380	cTP	Sha, Cvi-0
	S403C	probably damaging	0.838	substrate binding	Nok-2
AtTPI (*At1g21170*)	G14D	probably damaging	0.920	cTP	Can-0
	D201G	probably damaging	0.910	TPI domain	Bur-0, Nok-2
	S237F	possibly damaging	0.860	TPI domain	Est-1
AtPRK (*At1g32060*)	E208D	probably damaging	0.98	PRK family domain	Te-0
AtPGK (*At1g56190*)	S78G	possibly damaging	0.547	cTP	Cvi-0
	N99T	probably damaging	0.995	PGK domain	Te-0
	L124M	probably damaging	0.961	PGK domain	Te-0
	G334E	probably damaging	0.994	PGK domain	Sha

The detected non-synonymous substitutions of the investigated genes were tested for possible protein structure damaging substitutions using the PolyPhen2 server. This tool uses straightforward physical and comparative considerations for the prediction of possible impact of an amino acid substitution on the structure and function of a protein. Here we present only those substitutions which were predicted to have a possibly or probably damaging impact on protein structure.

Discussion

The analysis of genetic variation of Calvin cycle genes among accessions of *A. thaliana* as possible adaptation to different habitats revealed that genes involved in this particular primary metabolism are as variable as genes coding for secondary pathways. Despite high nucleotide diversity most of the genes are under purifying selection (Z-test) and possess negative Tajima's *D* values, a general feature of the *A. thaliana* genome reflecting the recent population expansion. Positive Tajima's *D* values and purifying selection were found in *AtGAPDH-a2* (Table 8), *AtFBPase* (Table 13), *AtSBPase* (Table 14), and *AtRPE* (Table 16). All 4 genes show separation in two major haplogroups rather than random substitutions. It appears possible that balancing selection drives the maintenance of two major haplogroups, with purifying selection acting within each haplogroup.

Several genes exhibit high numbers of non-synonymous substitutions (*AtRMT* (Table 4), *AtPGK* (Table 5), *AtGAPDH-b* (Table 9), *AtSBPase* (Table 14)). Some substitutions were predicted having damaging effect on protein structure (PolyPhen2). In those cases only single or two accessions exhibit the damaging substitution (Table 17), implying a general pattern of functional conservation. If affected enzymes are less effective or even non-functional, either duplicated genes or alternative pathways might functionally compensate the deficiency of the respective enzyme.

FORSYS Research Units for Systems Biology

SPONSORED BY THE Federal Ministry of Education and Research