Short thesis for the degree of Doctor of Philosophy (Ph.D.)

CHARACTERIZATION OF NOVEL PROTEIN PHOSPHATASES IN *DROSOPHILA* SPECIES

by Csaba Ádám

Supervisor: Viktor Dombrádi Ph.D, D.Sc.

UNIVERSITY OF DEBRECEN

Doctoral School of Molecular Medicine

Debrecen, 2011

Characterization of novel protein phosphatases in *Drosophila* species

By Csaba Ádám, M.Sc.

Supervisor: Viktor Dombrádi, Ph.D, D.Sc.

Doctoral School of Molecular Medicine, University of Debrecen

Head of the **Examination Committee**:

Members of the Examination Committee:

The Examination takes place at ……Medical and Health Science Center, University of Debrecen

Head of the **Defense Committee**:

Reviewers:

Members of the Defense Committee:

The Ph.D. Defense takes place at the Lecture Hall of the 1st Department of Medicine, Institute for Internal Medicine, Medical and Health Science Center, University of Debrecen

1. INTRODUCTION

1.1 Protein phosphorylation and dephosphorylation

The reversible phosphorylation of proteins is one of the most common post-translational modifications, which occurs in all the eukaryotic organisms. Phosphorylation reactions are carried out by protein kinases, so that they transfer the gamma phosphate group of ATP, less often of GTP, to the serine, threonine, and tyrosine side chains of proteins. The protein conformation is altered during the phosphorylation, whereby the function of a protein may be modified. Protein phosphorylation can be regarded as an effective regulatory tool only if it is possible to remove the phosphate group of a protein. The hydrolytic removal of a phosphate group is catalyzed by protein phosphatases.

1.2 The evolution of protein phosphatases

The significance of these two competing enzyme families is equivalent in propelling the phosphorylation-dephosphorylation based regulatory cycles, however, the phosphatases are second to the kinases in two respects. (1) According to the reaction mechanisms, kinases should act first by modifying the side chains in the nascent polypeptides. (2) Phosphatases must have evolved after the kinases, since in the absence of the phosphoprotein substrate they would have had no useful function. Some of the already existing hydrolytic enzymes adopted their catalytic pockets to accommodate the new substrates and slowly acquired more and more specificity.

1.3 The PPP enzyme family

The so called PhosphoProtein Phosphatase (PPP) enzymes developed from the bacterial diadenosine tetraphosphatases. PPP enzymes contain two metal ions in their active centre, which promote the hydrolysis of phosphate groups from Ser or Thr side chains of proteins. The PPP subfamily includes the Zn/Fe containing enzymes, PP1, PP2A, and PP2B (PPP1, PPP2, PPP3 according the new terminology) and the more recently discovered “novel” members of the PPP family that show a transition between the former three groups according to their structures and properties. The identification of novel enzymes was not possible by classical enzyme activity methods, but only with molecular cloning technique. These novel phosphatases were found in animals, plants and in numerous fungal species, where they regulate different physiological functions.

1.4 The *D. melanogaster* phosphoprotein phosphatase enzyme family

Drosophila melanogaster is a well known model organism of the molecular genetics. Based on a recent survey of the FlyBase we identified 19 genes coding for PPP catalytic subunits in this organism (Figure 1). According to their primary structures, the PPP enzymes can be divided into 5 subgroups (Table 1).

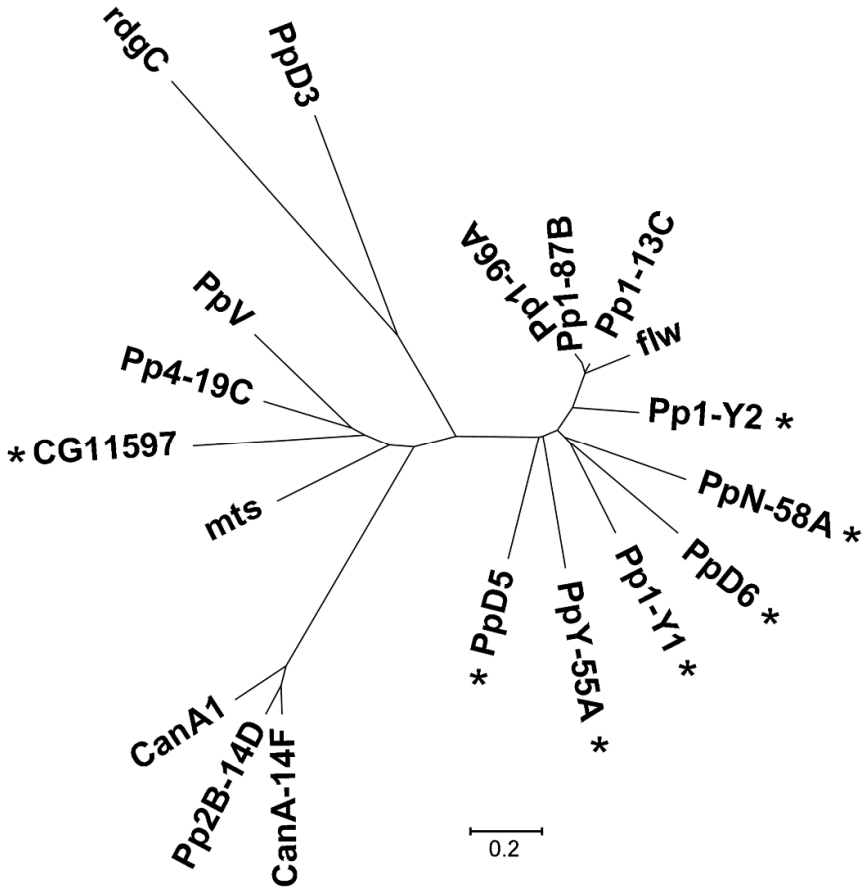

Figure 1. Phosphoprotein phosphatase (PPP) catalytic subunits of D. melanogaster
The bar indicates 20% amino acid substitutions per site. Drosophila-specific novel phosphatases are labelled by asterisks.

Classical and molecular genetic research suggested that the PPP genes have an important regulatory role in the fruitfly:

(1) The PPP1 family includes the *Pp1-13C*, *Pp1-87B*, *Pp1-96A* and *Pp1-9C* enzymes. The latter is affected in the *flapwing (flw)* mutant that exhibits aberrant flight muscle development. Mutations in the *Pp1-87B* gene cause a mitotic block in the anaphase. The inactivation of *Pp1-96A* does not result in any obvious phenotype. The *Pp1-13C* seems to be a redundant gene. In addition to the PPP1 enzymes, there are 6 novel members in the

subfamily: *PpY-55A, PpN-56A, PpD5, PpD6, Pp1-Y1,* and *Pp1-Y2*. Two of the genus-specific novel phosphatases, *PpY-55A*, and *PpN58A*, have been characterized previously in more detail. It was shown that both of them were exclusively expressed in the testis of male flies.

(2) The calcineurin/Pp2B/PPP3 Ca-regulated protein phosphatases are represented by 3 closely related isoforms. *CanA1* is a huge gene with 12 introns, while *Pp2B-14D*, and *CanA-14F* have no introns in their coding regions. *Pp2B-14D* has a role in the innate immune response and in the fertility of the females. *CanA-14F* has no similar functions.

(3) In *D. melanogaster* there are 4 type 2 phosphatases. In the absence of the *mts* gene, the organization of microtubules is abnormal. *Pp4-19C* was first described as PPX, and its role in the microtubule organization was identified by the analysis of the *centrosomes minus microtubules (cmm)* mutants. The accumulation of the protein in embryos implicated *PpV/Pp6* in zygotic transcription and cellularization. A new member of the family, *CG11597* was originally termed as *Pp4-like*. *CG11597* is a *Drosophila* specific novel phosphatase whose physiological significance has not been identified yet.

(4) *PpD3* still keeps the original name of a PCR product that led to its discovery. It is the ortholog of mammalian PPP5/Pp5. Although the regulation and the functions of the mammalian counterparts are well documented, much less is known about the *Drosophila* enzyme.

(5) *Retinal degeneration C (rdgC)* was discovered as a protein phosphatase protecting retina from light-induced degeneration.

1.5 The origin of the *D. melanogaster* PPP family

The large number of *D. melanogaster* PPP genes and the fact, that most of the PPP genes do not contain intron in the coding region, suggested that the enzyme family was underwent a significant transformation during evolution. According to recent estimates, about 17 genes duplicated in the *Drosophilidae* within 1 million years, while 5-11 genes originated in the melanogaster subgroup within the same period of time. The majority of the new genes is generated by tandem or dispersed duplications; in addition, retropositions, formation of chimeric genes, and *de novo* gene origination from noncoding sequences contribute to the extension of the gene repertoire. However, the gene number in a given family is also affected by deletion or degradation of some genes; a dynamic balance between gains and losses is a characteristic feature of genome evolution. The publication of the genome sequences of 12 *Drosophila* species (*D. melanogaster, D. simulans, D. sechellia, D. yakuba, D. erecta, D. ananassae, D. pseudoobscura, D. persimilis, D. willistoni, D. mojavensis, D. virilis, D.*

grimshawi) gave us an opportunity for the comparative analysis of PPP gene families. By comparing the nucleotide and the amino acid sequences in *Drosophilidae* we ventured on disclosing the evolution of the relatively small and strongly conservative gene family.

2. AIMS

The characterization of the novel PPP genes expression was not complete; we had only some microarray data, which suggested their male-specific expression. Additionally, we had only limited information about the evolution of novel and classical PPP genes.

To study the novel PPP genes function and their evolution, we formulated the following aims:

1) Characterization of novel PPP genes in *D. melanogaster*

- to study the expression of novel PPP genes during the lifecycle of *D. melanogaster*
- to study the expression patterns of novel PPP genes in the testis of *D. melanogaster* males
- to investigate the role of Y chromosome in the male-specific genes expression
- to study the functions of the *PpY-55A* and the *PpN58A* genes by RNAi method

2) Analysis the evolution of PPP genes in 12 *Drosophila* species

- to compare PPP genes in *Drosophila* and in other insect species
- to confirm and supplement the data of sequence databases
- to find novel PPP genes using sequence databases
- to investigate the expression of novel PPP genes
- to study the genomic rearrangements of phosphatases during their evolution
- to study the adaptation of PPP genes to genomic environment

3. MATERIALS AND METHODS

3.1 Materials

The majority of chemicals used in our experiments were obtained from Sigma-Aldrich. The PPP gene-specific oligonucleotide primers were purchased from Integrated DNA Technologies, Inc. *Escherichia coli* DH5α strain was purchased from Novagen. *Drosophila* strains were from Bloomington *Drosophila* Stock Center.

3.2 Methods

Classical genetic methods

Drosophila strains were cultivated on standard laboratory conditions and were selected by using well recognizable genetic markers.

D. melanogaster X/0 males and the XX/Y females were generated by using flies containing attached XY and attached XX chromosomes.

PpY-55A and *PpN58A* transgenic strains were generated with the help of Rozália Sinka and Péter Deák (HAS, BRC, Institute of Biochemistry, Szeged). Transgenic flies were driven by spermatocyte-specific driver.

Molecular biology methods

Plasmid DNA was isolated with QIAprep Spin Miniprep plasmid isolation kit according to the manufacturer's instructions. Genomic DNA was isolated from adult flies.

Total RNA was isolated from adult flies and from the testis of the males with Trizol Reagent. The DNA content of total RNA was removed by RNase-free DNase. The RNA was reverse-transcribed with M-MLV Reverse Transcriptase and oligo-dT primer.

Genomic DNA and cDNA fragments of different sizes were amplified by *Taq* polymerase enzyme. PCR products were purified on Microcon ultracel YM-100 columns for direct sequencing.

The *PpY-55A* and *PpN58A* genes were cloned into the second generation RNA interference pWIZ vector. The two genes were cloned into the vector in two copies forming an inverted repeat. These constructs were transformed into *E. coli* DH5α strains.

In situ hybridization was performed on the testes of adult Oregon-R males. Gene-specific PCR fragments were amplified from *Drosophila* genomic DNA template using either T7 (for sense probes) or T3 (for antisense probes) promoter tagged primers. The labelling

procedure was performed by DIG RNA Labeling Mix. Sense probes were used as negative controls.

The PCR products and the clones were sequenced at the BIOMI Ltd. in Gödöllő using PPP gene-specific primers or primer pairs.

Bioinformatic analysis

Sequence data were collected from the FlyBase the UCSC Genome Bioinformatics and the NCBI, databases. First, 19 known *D. melanogaster* PPP protein sequences were used as queries in blast for the homology search. Hits of the first round were used as queries in a second blastp search in order to confirm, and extend the sequence collection. The domain structure of the new hypothetical protein sequences were determined by SMART software.

Obvious mistakes of the databases (frame shifts, premature stops, insertions, and deletions) resulting in an aberrant protein that was not compatible with the conserved primary structures of the closely related PPP enzymes were corrected manually with the help of pDRAW32. If a PPP gene sequence was missing from the databases, their identification was performed by PCR reaction and DNA sequencing and the primers were designed based on the homologous gene sequences of the most closely related *Drosophila* species.

Protein sequences were compared with the pDRAW32, ClustalW, and the BioEdit software. Phylogenetic and molecular evolutionary calculations were conducted with MEGA 4.0 software. Orthologous relationships were deduced from the tree topology and were confirmed by microsyntheny analysis. The complex type 1 PPP subfamily was also analyzed by the multidimensional scaling method. In order to study the evolution of classical and novel phosphatases, R, dn and ds values were calculated with MEGA 4.0 software.

Gene localizations in the Muller elements were taken from FlyBase. Genetic rearrangements were examined by dot plots.

4. RESULTS

4.1 Expression pattern of novel PPP genes in *D. melanogaster*

Based on microarray data, the intronless retrogenes of *D. melanogaster* (*PpD5*, *PpD6*, *Pp1-Y1*, *Pp1-Y2 PpY-55A*, *PpN58A)* except for *CG11597* are male-specifically expressed. Two of the genus-specific novel phosphatases, *PpY-55A*, and *PpN58A*, have been characterized previously, so we focused our attention to other novel phosphatase. First the expression of the up till now less characterized phosphatase genes *PpD5*, *PpD6*, *Pp1-Y1*, *Pp1-Y2*, and *CG11597* was studied by RT-PCR during the morphogenesis of *D. melanogaster*. *CG11597* was expressed in all the developmental stages at a variable level. In contrast, *PpD5, PpD6*, *Pp1-Y1*, and *Pp1-Y2* were expressed only in the adult males. Since most of the studied PPP genes were expressed in males (except for *CG11597*), next we studied their expression in the testis.

The RT-PCR results collectively demonstrate that the tested novel phosphatases are selectively expressed in the male gonads, because they are not expressed in other parts of the body. Next we investigated the location of gene expression in the testis. It was performed by *in situ* hybridization method with digoxigenin-labeled antisense RNA probes. The previously characterized *PpY-55A* was used as a control gene; it is expressed in the distal (primary spermatocytes, nucleus of cyst cells) and the concave region (round spermatids) of the testis. The *PpD5* and the *Pp1-Y1* probes were detected mRNA in the same places, so they have a similar location to *PpY-55A*. *PpD6* exhibited a distinct spotted expression pattern, which corresponded to the distal ends of the elongating spermatid bundles. According to its location, *PpD6* is a post-meiotically expressed gene.

In order to study the male-specific expression of the genes, sterile X/0 males lacking Y chromosome and XX/Y females with Y chromosome were created. Gene expression was investigated by RT-PCR. The autosomal *PpD5*, *PpD6*, *PpY-55A*, and *PpN58A* gene products are present in the X/0 males. The Y-specific *Pp1-Y1*, and *Pp1-Y2* mRNAs are missing, as expected. In the XX/Y females, none of the tested phosphatases have been expressed. With these experiments we have proven the male-specific expression of the studied genes.

4.2 The investigation of *PpY-55A* and *PpN58A* gene functions

It is well known from the literature, that the above genes are expressed in the testis of the males. Since the creation of mutants was not possible by classical genetic methods, we studied the function of the genes using RNA interference (RNAi). Therefore we have generated *PpY-55A*, and *PpN58A* integrated strains, which were capable for RNAi. Since the genes are expressed in the testis, and *PpY-55A* is expressed in spermatocytes besides the cyst cells, the RNAi strains were crossed with spermatocyte-specific driver and homozygote strains were selected. The efficiency of RNA interference was proven by RT-PCR. We did not recognize phenotypic difference in the silenced strains in comparison to wild type. Our result supports the view that this group of phosphatases has overlapping functions allowing the compensation for the lack of one or the other member of the gene family. The other possible explanation is that gene silencing did not abolish totally the gene expression and there was enough gene product to maintain the wild type phenotype. Finally, it is important to mention, that the used pWIZ vector is not capable of germline expression, so our system could only work in somatic testis cells.

4.3 Evolution of PPP genes in *Drosophilidae*

The origin and evolution of *Drosophila* PPP genes (including novel phosphatases) was the less studied field of the literature. We used *D. melanogaster* as a gold standard for the comparison of the PPP catalytic subunits in 12 *Drosophila* species. First, we determined the minimal PPP toolkit comprising 7-8 enzymes that are present in most of the sequenced insect species (Table 1) (*Culex quinquefasciatus, Aedes aegypti, Anopheles gambiae, Tribolium castaneum, Apis mellifera, Nasonia vitripennis, Acyrthosiphon pisum, Pediculus humanus corporis*). The most of the PPP genes do not contain intron in the coding region, except for the *Pp2B* gene, suggesting that it was evolved by retroposition before the evolution of *Drosophila* species.

Next we investigated the presence of *D. melanogaster* PPP genes in the other 11 sequenced *Drosophilidae* species and determined their location in the Muller elements (the corresponding parts of the genome of *Drosophila* species) (Table 1). With a few exceptions, all of the PPP genes of *D. melanogaster* are also present in the genomes of the other 11 members of *Drosophilidae*, total PPP numbers fluctuate between 18 and 22 per species. The 7 of the 8 common insect phosphatases fall into 7 well-defined, distinct groups demonstrating a conserved structure and suggesting a conserved function for these enzymes. The phosphatases

of the non-*Drosophila* species form sub-branches that are well separated from the *Drosophila* proteins supporting a common origin but independent evolution.

4.4 The PPP complement of *Drosophilidae*

Strikingly, in addition to the above phosphatase repertoire, the members of *Drosophilidae* acquired no less than 15 new, dynamically changing PPP family members (Table 1, Figure 2).

Table 1. Localization of PPP genes in Muller elements of 12 Drosophila species

Classification	**Gene name, synonym[a]/Species[b]**	Dmel	Dsim	Dsec	Dyak	Dere	Dana	Dpse	Dper	Dwil	Dmoj	Dvir	Dgri
Type 1 or PPP1	***Pp1α-96A, Pp1-96A***	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**
	Pp1-87B	***E***	***E***	***E***	***E***	***E***	***E***	***E***	***E***	***E***	***E***	***E***	***E***
	Pp1-13C	***A***	***A***	***A***	***A***	***A***	***A***	***A***	***A***	***A***	***A***	***A***	***A***
	flw, Pp1-9C	**A**	**A**	**A**	**A**	**A**	**A**	**A**	**A**	**A**	**A**	**A**	**A**
	Pp1-Y1	***Y***	***Y***	***Y***	***Y***	***Y***	***Y***	***C***	***C***	***C***		***C***	***C***
	Pp1-Y2	***Y***	***Y***	***Y***	***Y***	***Y***	***Y***	***C***	***C***	***C***	***C***	***C***	***C***
	PpD5	***C***	***C***	***C***	***C***	***C***	***C***	***C***	***C***		***C***	***C***	
	PpD5+							**D**	**D**		***D***	***D***	***D***
	PpD6	***B***	***B***	***B***	***B***	***B***	***Y***	***C***	***C***	***C***	***C***	***C***	***C***
	PpD6+							***C***	***C***	***C***		***C***	
	PpN58A	***C***	***C***	***C***	***C***	***C***	***C***	***C***	***C***	***C***	***C***	***C***	***C***
	PpY-55A	***C***	***C***	***C***	***C***	***C***	***C***	***A***	***A***	**C**	**C**	**C**	
	PpY+							***A***	***A***				
Calcineurin or Pp2B or PPP3	***CanA1***	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**
	Pp2B-14D	***A***	***A***	***A***	***A***	***A***	***A***	***A***	***A***	***A***	***A***	***A***	***A***
	CanA-14F	***A***	***A***	***A***	***A***	***A***	***A***	***A***	***A***	***A***	***A***	***A***	***A***
Type 2 or PPP2-4-6	***mts, Pp2A***	**B**	**B**	**B**	**B**	**B**	**B**	**B**	**B**	**B**	**B**	**B**	**B**
	Pp4-19C	**A**	**A**	**A**	**A**	**A**	**A**	**A**	**A**	**A**	**A**	**A**	**A**
	Pp4+							***B***	***B***				
	PpV, Pp6	**A**	**A**	**A**	**A**	**A**	**A**	**A**	**A**	**A**	**A**	**A**	**A**
	Pp6+										***E***	***E***	***E***
	CG11597	***D***	***D***	***D***	***D***	***D***	***D***						
PPP5	***PpD3, Pp5***	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**
PPP7	***rdgC, Pp7***	**D**	**D**	**D**	**D**	**D**	**D**	**D**	**D**	**D**	**D**	**D**	**D**
PPP	Teljes szám	**19**	**19**	**19**	**19**	**19**	**19**	**22**	**22**	**18**	**19**	**21**	**18**

[a] The names of the genes that constitute the minimal PPP toolkit are written in bold face.
[b] Species names are abbreviated as follows: Dmel: D. melanogaster, Dsim: D. simulans, Dsec: D. sechellia, Dyak: D. yakuba, Dere: D. erecta, Dana: D. ananassae, Dpse: D. pseudoobscura, Dper: D. persimilis, Dwil: D. willistoni, Dmoj: D. mojavensis, Dvir: D. virilis, Dgri: D. grimshawi.
The localization of intronless genes is highlighted in underlined italics.
Empty spaces stand for a missing gene.

Duplications are supposed to be the most effective tools for extending gene repertoire. Our data support the thesis that both Pp2B-14D and CanA-14F are the intronless derivatives of CanA1 in all *Drosophila* species. The comparison of the phosphatases in 8 insects, and especially in 12 *Drosophila* species suggests that an intron containing ancient form of *Pp1-96A* was the parent of the intronless type 1 phosphatases. *Pp1-87B* evolved directly from *Pp1-96A* by retroposition. In addition to the duplication of classical phosphatase genes, the PPP set of *Drosophilidae* was substantially expanded by the appearance of more retrogenes (Table 1, Figure 2). The novel type 1 phosphatases, *Pp1-Y1*, *Pp1-Y2*, *PpD5*, *PpD6*, *PpN58A*, and *PpY-55A*, that were originally discovered in *D. melanogaster*, are present in all of the 12 *Drosophila* species, but are absent from other insects. As a matter of fact, their orthologs cannot be found in any other living organisms, thus they can be classified as *Drosophila* specific phosphatases.

Retropositions created additional new PPP enzymes (PpD5+, PpD6+, PpY+, Pp4+, Pp6+), that do not exist in *D. melanogaster* (Table 1, Figure 2). The new genes were inserted into new chromosomal places; their ancient paralogs preserved the original location. According to our sequence comparisons it is likely that *PpD6+* and *PpD5+* were derived from *PpD6* and *PpD5*, respectively. The *PpY+* and *Pp4+* retrogenes were in identified two of the closely related representatives of the obscura group, *D. pseudoobscura*, and *D. persimilis*. The similarity between the *PpY-55A* and *PpY+* sequences, as well as the close localization of the two genes in the physical map of Muller element A suggest that the latter was generated from the former by replicative transposition. Likewise, sequence analysis of the *Pp4-19C* and *Pp4+* pairs reveals that the retroposition of the intron containing *Pp4-19C* produced the intronless *Pp4+* gene copy.

Although the new PPP genes code for full length proteins that contain the critical residues which are necessary for phosphatase activity, without experimentation it would be rather difficult to predict if they were transcribed, or were degraded to an inactive pseudogene status. In order to prove it we investigated the expression novel PPP genes (including the known genes and those identified in the present sudy) and their paralogs during the evolution of the *Drosophila* species. We have chosen 4 distantly related *Drosophila* species for our experiments. We used *D. ananassae* from the melanogaster group, *D. pseudoobscura* from the obscura group, *D. willistoni* from the willistoni group, and *D. virilis* from the virilis group. We have proven that all of the new genes are functionally active retrogenes and most of them have a strict male-specific expression pattern. The only exception was *CG11597*, which was expressed in both sexes.

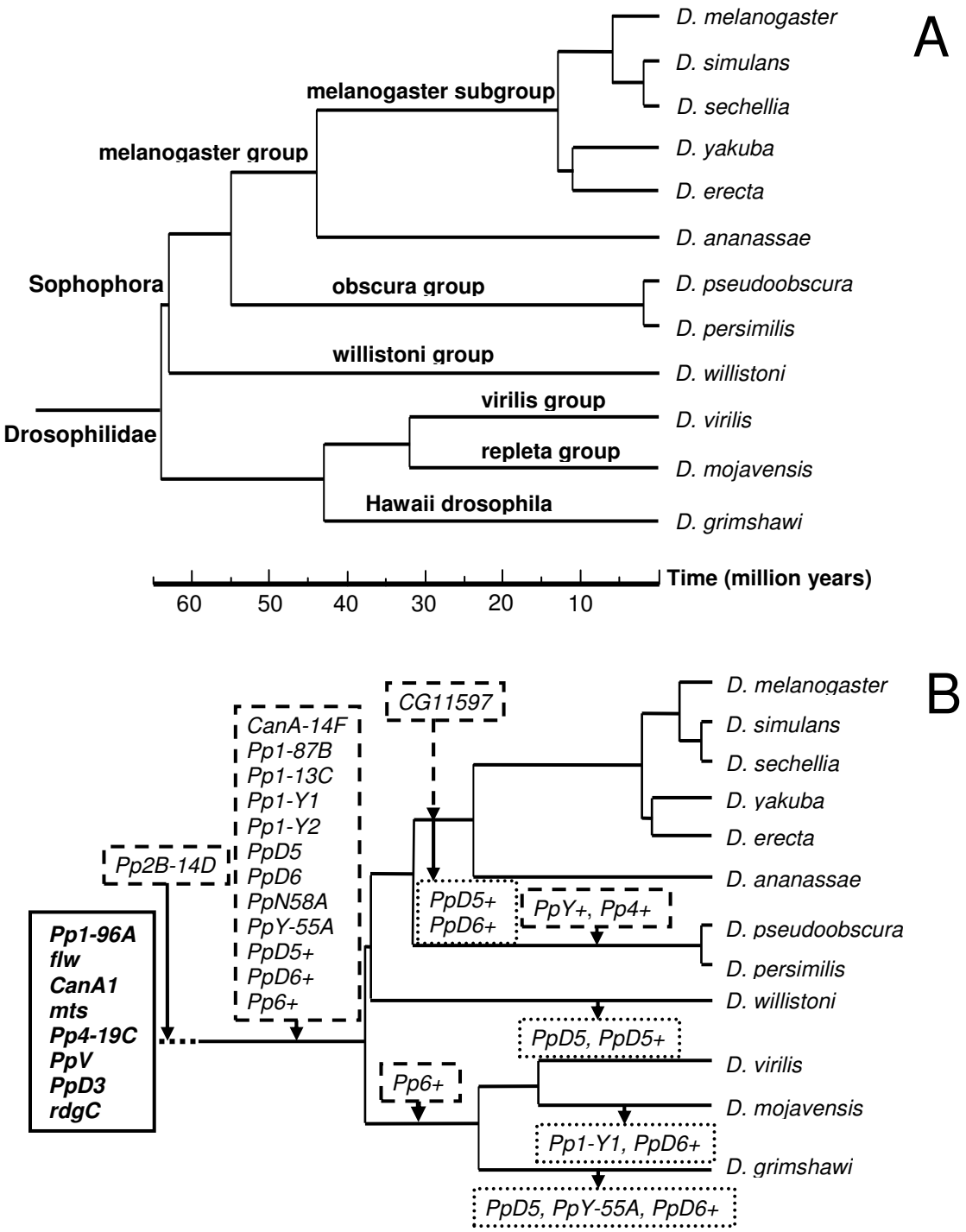

Figure 2. Evolution of PPP genes in Drosophilidae

A part *of the figure shows the evolution of 12 Drosophila species. The main subspecies, groups and subgroups are labeled.*

B part *depicts the evolution of PPP genes. The basic PPP set of insects is in a black box. Gained genes are in broken line boxes and lost genes are in dotted line boxes. The broken line indicates that the Pp2B-14D gene was gained in Culex quinquefasciatus, Aedes aegypti, and Anopheles gambiae insect species.*

4.5 The fast evolution of male-specific PPP genes in *Drosophila* species

The evolution of genes with male-specific functions is well known to be faster (the ratio of non-synonymous to synonymous substitutions is higher) than that of the non-specialized genes. We wanted to investigate whether our phosphatases also follow this rule. Therefore we analyzed the sequences encoding catalytic domains of the classical and novel, mainly male-specific phosphatases in 12 *Drosophila* species. The total transition/transversion ratio (R), the number of synonymous (ds) and non-synonymous (dn) substitutions and the dn/ds ratio was calculated. Relative to the coding sequences of the unbiased metabolic phosphatases, the male-specific PPP enzymes exhibit a significantly higher dn/ds substitution ratio, in agreement with the accelerated evolution of the male-biased genes. In addition, our results proved the "out of testis" hypothesis; while the older PPP retrogenes developed a differential expression pattern and gained specialized functions, the youngest PPP retrogenes are still male specific.

4.6 The changes of PPP gene numbers and locations

Besides the gain of genes the gene losses, the changes in the chromosomal localizations and gene structures also contributed to the evolution of the PPP family in *Drosophilidae*. The absence of a gene from the database may be attributed to the intrinsic shortcoming of the shotgun sequencing strategy. To overcome this problem, we used PCR for the amplification and identification of the missing genes (Table 1). For reference, the localizations of PPP genes in the Muller elements of *D. melanogaster* were used. The data collectively demonstrate that the PPP genes keep a well conserved position and orientation in the members of the melanogaster subgroup. On the other hand, many PPP genes, and first of all the novel type 1 phosphatase genes, frequently change positions (Table 1). The role of gene rearrangements in the PPP gene evolution is demonstrated by the following 3 examples. (1) The *PpY-55A* gene evolved via an intron containing intermediate that was preserved in the members of the repleta, virilis, and willistoni groups. The *PpY-55A* gene was lost from *D. grimshawi*. (2) In *D. pseudoobscura* and *D. persimilis*, the *PpD5+* gene has a unique, 60 bp intron in the coding region. This gene was deleted from the members of the melanogaster and willistoni groups, but its intronless version was established in all of the ancient *Drosophila* subspecies. (3) The *PpD6, PpD6+, Pp1-Y1, Pp1-Y2* genes are next to each other in the same orientation in the gene cluster of *D. virilis*. This gene cluster was disintegrated in the melanogaster group. *PpD6+* was eliminated in all of the group members. In *D. ananassae* one, half of the cluster, *Pp1-Y2* and *PpD6* moved together into the Y chromosome. In the

common ancestor of the melanogaster subgroup *PpD6* landed in element B while *Pp1-Y1* and *Pp1-Y2* arrived to heterochromatic areas of Y (Table 1).

4.7 Adaptation of the PPP genes to the genomic environment

It has been reported previously that *Drosophila* genes residing in a heterochromatic environment have lower G-C content than their paralogs in euchromatic regions. By extending the idea of gene adaptation we analyzed if the movements of the PPP genes between the autosomal chromosomes and the heterochromatin rich Y chromosome affected their base composition. The average G-C content of the PPP coding regions varies between 50-60%, except for *D. willistoni* that has a lower level of these nucleotides. *D. willistoni* has a lower G-C content because it prefers T over C in the codons of certain amino acids. First, we studied the G-C content changes of the *PpD6, PpD6+, Pp1-Y1, Pp1-Y2* gene cluster during its rearrangements. In accord with the expectation, the G-C content of the phosphatase cluster members is close to average when they are in euchromatic environment, but the ratio becomes significantly reduced when they move to Y chromosome. The G-C content of all known *Drosophila* genes that have been relocated between the Y and other chromosomes was also compared. We have found that the G-C content of other genes relocated to the Y chromosome was also reduced.

4.8 Conclusions

In conclusion, the relatively small PPP gene family reflects many colorful molecular events of evolution. Our work demonstrates that retropositions, tandem duplications, deletions and relocations have steadily modified the PPP repertoire of fruit flies. From this respect *Drosophilidae* is an especially resourceful organism as it accumulated one of the largest PPP complements in the animal kingdom. The dynamic alterations including the changes in the numbers, structures, orientation, and chromosomal localization of PPP genes contributed to the genetic diversity in *Drosophilidae*.

5. SUMMARY

Protein phosphorylation and dephosphorylation are common gears of regulation in all eukaryotic organisms. The group of the so-called phosphoprotein phosphatases (PPP) removes the phosphate from the Ser and Thr residues of proteins. The PPP family comprises classical members that were identified by biochemical assays and novel enzymes that were discovered by molecular biology or genetic approaches.

In the genome of *Drosophila melanogaster* we have identified 19 phosphoprotein phosphatase (PPP) catalytic subunit coding genes. Seven of the novel members of the gene family turned out to be *Drosophila*-specific. *CG11597* is a recently evolved gene that is expressed during all stages of morphogenesis in *D. melanogaster*. In contrast, transcription of the *PpD5*, *PpD6*, *Pp1-Y1*, and *Pp1-Y2* genes is restricted to the pupa and imago developmental stages and to the testes of the males, just as that of the previously characterized *PpY-55A* and *PpN58A*. The mRNA of *PpD5*, *Pp1-Y1*, and *PpY-55A* were detected in the developing cysts by *in situ* hybridization, in contrast with the *PpD6* transcript that was found in the distal ends of elongating spermatids. The localization suggests that *PpD6* is one of the few post-meiotically transcribed genes in *D. melanogaster*.

Based on the genome sequences of 12 *Drosophila* species we traced the evolution of the PPP catalytic subunits. We noted a substantial expansion of the gene family. We concluded that the 18-22 PPP genes of *Drosophilidae* were generated from a core set of 8 indispensable phosphatases that are present in most of the insects. Retropositions followed by local gene duplications extended the basic phosphatase repertoire, and sporadic gene losses contributed to the species specific variations in the PPP complement. During the course of these studies we identified 5 up till now uncharacterized phosphatase retrogenes: *PpY+, PpD5+, PpD6+, Pp4+,* and *Pp6+*, which are found only in some ancient *Drosophila* species. We demonstrated that all of these new PPP genes exhibit a distinct male specific expression. Our data support the "out of testis" hypothesis suggesting that the new functional retrogenes are preferentially transcribed in the male gonads. We have also proved that the sequence of novel, male-specific phosphatases changed more rapidly than that of the classical phosphatases, thus our results support the "faster male" hypothesis. In addition to the changes in gene numbers, the intron-exon structure and the chromosomal localization of several PPP retrogenes was also altered during evolution. The G-C content of the coding regions decreased when a gene moved into the heterochromatic region of Y chromosome. In conclusion, the PPP enzyme family exemplifies the various types of dynamic genome rearrangements that accompany the molecular evolution of the novel retrogenes in *Drosophilidae.*

6. PUBLICATIONS

UNIVERSITY AND NATIONAL LIBRARY UNIVERSITY OF DEBRECEN
KENÉZY LIFE SCIENCES LIBRARY

Register Number: DEENKÉTK /171/2011.
Item Number:
Subject: Ph.D. List of Publications

Candidate: Csaba Ádám
Neptun ID: SKORMA
Doctoral School: Orvosi Vegytani Intézet

List of publications related to the dissertation

1. Miskei, M., **Ádám, C.**, Kovács, L., Karányi, Z., Dombrádi, V.: Molecular Evolution of Phosphoprotein Phosphatases in Drosophila.
PLoS ONE. 6 (7), e22218, 2011.
DOI: http://dx.doi.org/10.1371/journal.pone.0022218
IF:4.411 (2010)

2. **Ádám, C.**, Henn, L., Miskei, M., Erdélyi, M., Friedrich, P., Dombrádi, V.: Conservation of male-specific expression of novel phosphoprotein phosphatases in Drosophila.
Dev. Genes Evol. 220 (3-4), 123-128, 2010.
DOI: http://dx.doi.org/10.1007/s00427-010-0332-6
IF:2.008

The Candidate's publication data submitted to the Publication Database of the University of Debrecen have been validated by Kenezy Life Sciences Library on the basis of Web of Science, Scopus and Journal Citation Report (Impact Factor) databases.

26 July, 2011

DEBRECENI EGYETEM KENÉZY ÉLETTUDOMÁNYI KÖNYVTÁRA EGYETEMI ÉS NEMZETI KÖNYVTÁR

H-4032 Debrecen, Egyetem tér 1. e-mail: publikaciok@lib.unideb.hu

Posters related to the thesis:

Ádám Csaba, Miskei Márton, Nagy Olga, Pál Margit, Deák Péter, Friedrich Péter, Dombrádi Viktor (2007) Az új típusú protein foszfatázok vizsgálata *Drosophila* fajokban**.** Magyar Biokémiai Egyesület 2007. évi vándorgyűlése**.** Debrecen.

Ádám Csaba, Miskei Márton, Nagy Olga, Pál Margit, Deák Péter, Friedrich Péter, Dombrádi Viktor (2008) Új típusú protein foszfatázok expressziója és evolúciója. 38. Membrán Transzport Konferencia. Sümeg.

Ádám Csaba, Pál Margit, Deák Péter, Friedrich Péter, Dombrádi Viktor (2008) Transzgénikus RNS interferencia módszer a PPY és PPN új típusú protein foszfatázok funkciójának tanulmányozására *Drosophila melanogaster*-ben**.** Magyar Biokémiai Egyesület 2008. évi vándorgyűlése**.** Szeged.

Ádám Csaba, Henn László, Erdélyi Miklós, Friedrich Péter, Dombrádi Viktor (2009) Új típusú foszfoprotein foszfatázok tesztisz specifikus expressziója. VIII. Magyar Genetikai Kongresszus, XV. Sejt- és Fejlõdésbiológiai Napok. Nyíregyháza.

Ádám Csaba, Henn László, Erdélyi Miklós, Friedrich Péter, Dombrádi Viktor (2009) Új típusú foszfoprotein foszfatázok expressziójának vizsgálata *Drosophila melanogaster*-ben. 39. Membrán-Transzport Konferencia. Sümeg.

Ádám Csaba, Miskei Márton, Dombrádi Viktor (2009) A foszfoprotein foszfatázok evolúciója *Drosophila* fajokban. Magyar Biokémiai Egyesület 2009. évi vándorgyűlése. Budapest.