



# Divergent evolution and purifying selection of the Type 2 diabetes gene sequences in *Drosophila*: a phylogenomic study

Manoj Kumar Gupta<sup>1</sup> · Ramakrishna Vadde<sup>1</sup>

Received: 30 November 2019 / Accepted: 12 August 2020 / Published online: 17 August 2020  
© Springer Nature Switzerland AG 2020

## Abstract

The recently developed phylogenomic approach provides a unique way to identify disease risk or protective allele in any organism. While risk alleles evolve mostly under purifying selection, protective alleles are evolving either under balancing or positive selection. Owing to insufficient information, authors employed the phylogenomic approach to detect the nature of selection acting on type 2 diabetes (T2D) genes in *Drosophila* genus using various models of CODEML utility of PAML. The obtained result revealed that T2D gene sequences are evolving under purifying selection. However, only a few sites in membrane proteins encoded via *CG8051*, *ZnT35C*, and *kar*, are significantly evolving under positive selection under specific scenarios, which might be because of positive or adaptive evolution in response to changing niche, diet or other factors. In the near future, this information will be highly useful in the field of evolutionary medicine and the drug discovery process.

**Keywords** Type 2 diabetes · Natural selection · *Drosophila* · Phylogenomics

## Introduction

Diabetes mellitus (DM) is a polygenic disease. Clinically it is characterized by hyperglycemia, polyuria (frequent urination), polyphagia (hunger), polydipsia (thirst), and loss of weight. DM is mainly of two types, namely, type 1 diabetes (T1D) & type 2 diabetes (T2D). T1D is clinically characterized via autoimmune destruction of insulin-producing pancreatic  $\beta$  cells; if not treated early results in absolute insulin deficiency. T2D is characterized via resistance towards the action of insulin as well as an incapability to produce adequate levels of insulin for overcoming 'insulin resistance' (Walker and Colledge 2013). While T1D is an autoimmune disease, obesity is a major risk factor for causing T2D along with various other genetics as well as environmental factors (Thirlaway and Davies 2001; Baynest 2015; Skyler et al. 2017). As T2D is a complex disorder, there are

still numerous debates on the actual cause, mechanism, and treatment associated with T2D. Thus, interest to understand the mechanism as well as to find a possible therapeutic for T2D with minimum side effects has revolutionized the field of diabetic research. Researchers are implementing several new technologies, like nanotechnology, statins, and gene therapy, for the treatment of T2D. Nevertheless, these new technologies, along with the traditional medicinal approach, are also reported to have certain side effects. For instance, the consumption of nanoparticles may be toxic or harmful (Tiwari 2015). Thus, there is always an urge to detect key gene(s) and its site(s) that play a significant role in the development of T2D, which in turn may function as a plausible therapeutic target towards the treatment of T2D.

The recently developed phylogenomic approach provides a unique way to understand how natural selection has shaped the genetic diversity of any organism. The phylogenomic approach also provides us a unique way to identify disease risk or protective allele in any organism. While risk alleles evolve mostly under purifying selection, protective alleles are evolving either under balancing or positive selection (Gupta and Vadde 2019a). However, when an individual with protective alleles migrate to the contrasting environment, the protective alleles may turn into risk factor and causes diseases (Gupta and Vadde 2019a). Hence, there is an urgent requirement to identify disease-specific protective or

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s10709-020-00101-7>) contains supplementary material, which is available to authorized users.

✉ Ramakrishna Vadde  
vrkrishna70@gmail.com;  
vrkbtbi@yogivemanauniversity.ac.in

<sup>1</sup> Department of Biotechnology & Bioinformatics, Yogi Vemana University, Kadapa, Andhra Pradesh 516005, India

risk alleles (Gupta and Vadde 2019a). This can be achieved by employing various principles of evolution (Grunspan et al. 2017). Through natural selection, it's easy to understand the processes associated with the adaptation of an organism to its environment through selectively reproducing changes in its genotype or genetic constitution (König 2001; Gupta and Vadde 2019a). By performing a comparative study of gene(s) sequences of closely related species, we can quickly determine the evolutionary relationship for a particular phenotype (e.g., T2D) (Fischman et al. 2011). Since the protein-coding region of any gene is highly conserved throughout evolution, estimation of evolutionary pressure on the protein-coding region provides more significant results in comparison to the non-coding region (Yang 2006). Evolutionary constraints on proteins across divergent lineages are generally estimated via the ratio of substitution rates at nonsynonymous (dN) and synonymous sites (dS) in the protein-coding regions ( $\omega = dN/dS$ ) (Yang 2006). Using the synonymous polymorphisms as a proxy of neutral diversity, one can easily predict whether nonsynonymous polymorphisms have been favored or hindered by natural selection (Yang 2006). In the case of neutral evolving genes, the rate of fixation of synonymous and nonsynonymous mutations will be the same ( $\omega = 1$ ) (Yang 2006). In the case of negative (purifying) selection, the nonsynonymous mutation is not favored via natural selection. It thus is eliminated, causing the rate of fixation of synonymous mutation to be higher than the nonsynonymous rate ( $\omega < 1$ ) (Yang 2006). In the case of positive (adaptive) selection, the nonsynonymous mutation is favored by positive selection, causing the rate of fixation of nonsynonymous mutation to be higher than the synonymous rate ( $\omega > 1$ ) (Yang 2006). Thus,  $\omega$  value can be employed extensively to understand evolutionary rates of genes, identify least or most conserved genes and also detect genes that may have undergone periods of adaptive (or positive) evolution (Kosiol et al. 2008). For instance, in parasite genomes,  $\omega$  value enables us to detect rapidly evolving genes in the “evolutionary arms race” against the host's immune system (Yang et al. 2003; Lefébure and Stanhope 2009).

Earlier phylogenomics approaches were utilized to understand antibiotic resistance and pathogen evolution as well as detecting the origins of emerging diseases, for instance, the origin of HIV1 in chimpanzees in Central Africa (Nesse and Stearns 2008). Phylogenomics approaches have also been employed in cancer treatment and research. Cell lines segregate under the influence of mutations, and the genetic differences make it possible to trace the original wild type sequence. Two tumors with identical histological features may have, unlike proteomic signatures, that, in turn, will help us to understand the degree of cellular differentiation. Whether the tumor has developed from the same line of cell or have different origin can also be detected via phylogenomics approaches (Nesse and Stearns 2008). Recently,

Al-Daghri and the team have reported that *G6PC2* genes are evolving under positive selection in mammal and play a significant role in causing T2D (Al-Daghri et al. 2017). *G6PC2* encodes a glucose-6-phosphatase catalytic subunit isoform that catalyzes the hydrolysis of glucose-6-phosphate for producing glucose as well as inorganic phosphate in the endoplasmic reticulum lumen (Pound et al. 2013). In another study, Klimentidis and the team identified three genes, namely, *IGF2BP2*, *WFS1*, and *SLC30A8*, that are under positive selection and also responsible for increasing the risk of T2D in East Asians and Sub-Saharan Africans human population (Klimentidis et al. 2011). In contrast, few studies reported that high milk consumption in Europe caused a positive selection of protective variants in milk-consuming populations, which might explain the low prevalence of T2D in Europeans (Ségurel et al. 2013). Thus, there is always a quest to understand how the evolutionary process has shaped the genetic makeup of the T2D gene in different organisms/populations and detect and characterize positively selected genes and their sites that may be responsible for either causing T2D or providing protection against T2D.

Because of high conservation between humans and flies at both physiological and molecular levels, *Drosophila* has served as the best useful model organism for studying a variety of human traits and diseases, including T2D (Musselman et al. 2011; Alfa and Kim 2016; Graham and Pick 2017). Till date numerous studies have been performed for understanding the influence of natural selection on the genetic diversity of numerous gene families, for example, male reproductive genes (Ahmed-Braimah et al. 2017), olfactory and gustatory receptors (Gardiner Anastasia et al. 2009) and immune genes (Hill et al. 2019) in *Drosophila*. However, no phylogenetic studies have been undertaken on the T2D gene of *Drosophila*. Thus, there is always a debate about how evolution has shaped the genetic diversity in the T2D genes of the *Drosophila* genus. Hence, for the first time, authors have utilized aligned protein-coding sequence files of T2D genes of 12 species of *Drosophila* available in the FlyBase R6.14 database (<https://flybase.org/>) for detecting nature of selection acting on T2D gene in *Drosophila*. In the near future, information obtained from the present study will help us in understanding the mechanism associated with T2D in humans, which in turn may be useful in the field of evolutionary medicine, as well as in the drug discovery process.

## Materials and methods

### Data retrieval and preprocessing

Aligned coding sequence (CDS) file of the longest isoform of every *D. melanogaster*'s genes that are present in all 12 species of *Drosophila* (*D. melanogaster*, *D. sechellia*, *D.*

*erecta*, *D. simulans*, *D. yakuba*, *D. ananassae*, *D. persimilis*, *D. pseudoobscura*, *D. willistoni*, *D. virilis*, *D. grimshawi* and *D. mojavensis*) were downloaded from FlyBase R6.14 databases ([ftp://ftp.flybase.net/genomes/12\\_species\\_analyses/clark\\_eisen/alignments/](ftp://ftp.flybase.net/genomes/12_species_analyses/clark_eisen/alignments/)). For maintaining consistency, CDS file lacking genes sequence of any of the 12 species were discarded. Further, a stop codon from each aligned sequenced file was discarded via bppsuite (<https://github.com/BioPP/bppsuite>).

## Orthologs search

FlyBase' gene IDs present in all 12 species of *Drosophila* were subjected to the DIOPT (“*Drosophila* RNAi Screening Center Integrative Ortholog Prediction Tool”) Diseases and Traits (<https://www.flyrnai.org/diopt-dist>). Tool for identifying high-confidence orthologs of human T2D genes in *Drosophila melanogaster* (Hu et al. 2011). DIOPT-DIST is comprised of zebrafish, human, fly, yeast, mouse, and worm ortholog predictions made by several existing tools, like, HomoloGene, Inparanoid, Ensembl Compara, Isobase, OMA, Phylome, RoundUp, TreeFam, and orthoMCL. Based on this information, DIOPT-DIST estimates simple scores that represent the number of tools that support a given orthologous gene-pair relationship, as well as a weighted score based on functional assessment using high quality GO molecular function annotation of all fly-human orthologous pairs predicted by each tool. These scores are represented as low, moderate, and high, where low denotes “least significant orthologous gene-pair relationship” and high denotes “highly significant orthologous gene-pair relationship” (Hu et al. 2011).

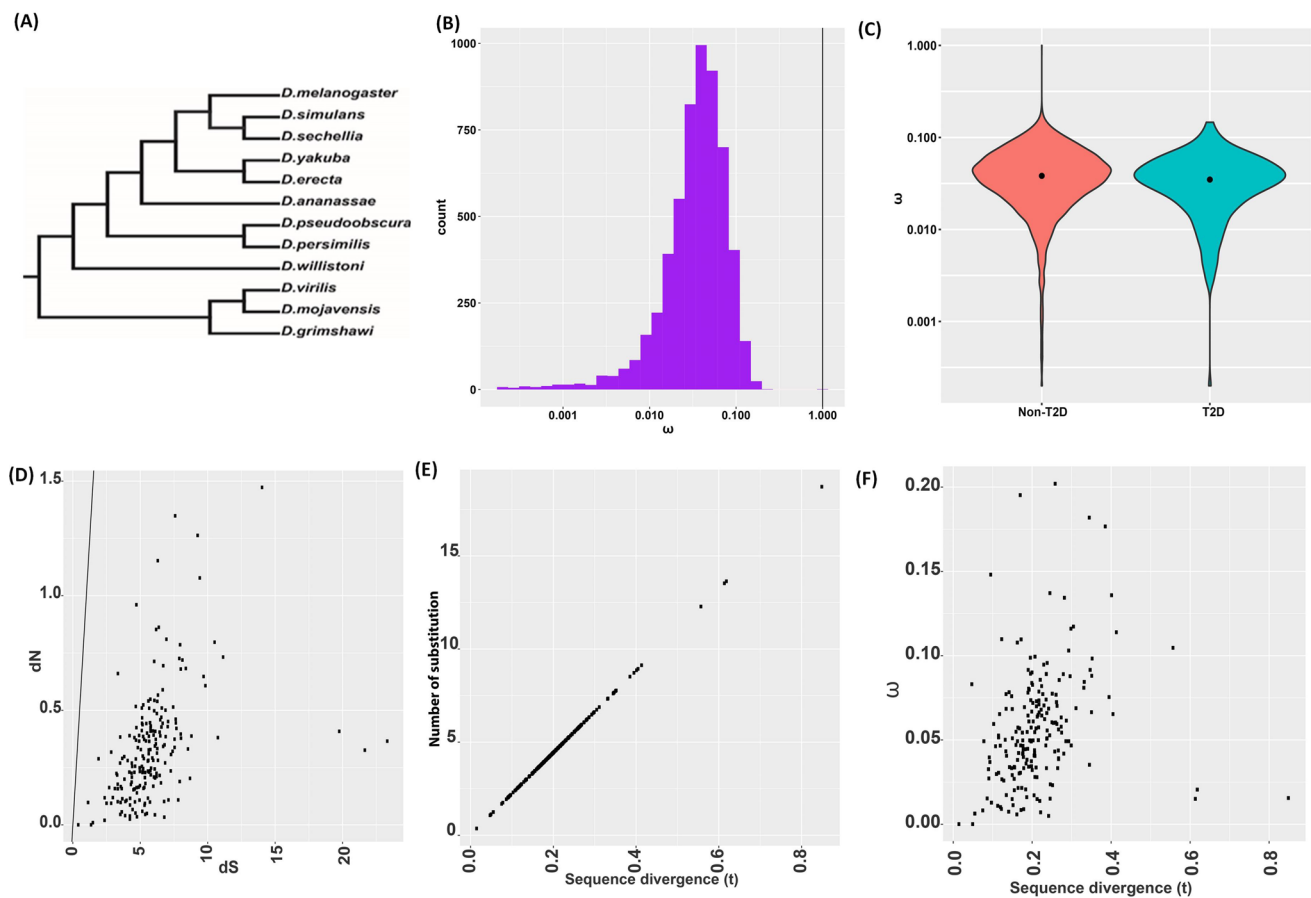
## Identification of selection pressure acting on T2D genes in the *Drosophila* genus

Initially,  $\omega$  value of every gene was calculated individually by using M0 (one ratio or neutral) model implemented in Phylogenetics Analysis by Maximum Likelihood (PAML) package v4.9 (Yang 2007). M0 is the simplest model of PAML and presume identical  $\omega$  value for all branches in a phylogenetic tree and across all sites (Swanson et al. 2003; Lynn et al. 2005). For estimating  $\omega$  value, the M0 model utilized the F3x4 model and gene tree of 12 species generated by the *Drosophila* Genome Consortium (Fig. 1A) (*Drosophila* 12 Genomes Consortium 2007). Later, quantile regression analysis was carried out between  $\omega$  value of T2D and non-T2D genes using the QUANTREG package of R (Koenker et al. 2018; Team 2014) to detect the strength of selective pressure that is responsible for shaping the function of T2D across entire *Drosophila* genus. Results having a  $p$ -value < 0.05 was considered significant.

Though significant, the result obtained from the M0 model reveals only the global scenario across any genus. However, earlier several studies have also suggested that the evolution shapes each branch of the phylogeny distinctly because the number of rates of nonsynonymous and synonymous substitutions varies across a sequence (Wong et al. 2008). Thus, in the present study, “Branch-site models” were also employed for detecting selection pressure that acts distinctly on T2D genes of each species of *Drosophila* (Farfán et al. 2009). “Branch-site models” allow  $\omega$  value to vary both amongst sites as well as lineages. Subsequently, quantile regression analysis between the  $\omega$  value of T2D and non-T2D genes of each species was performed separately to detect the strength of selective pressure. Results having  $p$ -value < 0.05 is considered significant.

## Detecting positively selected sites in T2D genes of *Drosophila* genus

Since genes and their sites that are evolving under positive selection are beneficial, there is always a quest to detect them (Wagner 2007). Earlier studies have also reported that few sites in genes that are evolving under purifying selection may also occasionally experience adaptive change (Yang and Bielawski 2000). Such sites point to functionally important gene's regions. Hence, they are of potential interest to protein engineers who alter proteins to produce new functions. By considering the above information, in the present study, T2D genes and their sites that are evolving under positive selection were detected using “fixed-sites models”, namely M7 and M8 (Yang and Swanson 2002). M7 allows 10 sites class following a beta-distribution of sites with  $\omega$  value ranging between 0 and 1. M8 model is similar to M7 models, except there is an additional 11th site class with  $\omega > 1$  (positive selection allowed). For estimating significance in terms of a  $p$ -value, a likelihood-ratio test (LRT) was employed. LRT ( $2\Delta\ell$ ) is computed as  $2(\ell_1 - \ell_0)$ , where  $\ell_1$  is the log-likelihood (LL) of the model representing the alternative hypothesis (M8) and  $\ell_0$  is the LL of the model representing the null hypothesis (M7). LRT statistic approximately follows a chi-square distribution.  $p$ -value obtained was further adjusted via “FDR” (false discovery rate) function in the R package. T2D gene and its sites with FDR value < 0.05 were considered to be significantly under positive selection. Later, Bayes empirical Bayes (BEB) methods available in PAML4.9 were employed to detect if any positive selection episodes had affected any specific amino acid sites in the protein encoded via genes (Liu et al. 2013; Teng et al. 2017). These T2D genes having positively selected sites are henceforth known as key genes.



**Fig. 1** **A** Gene tree of 12 *Drosophila* species generated by Drosophila Genome Consortium. **B** Global distribution of  $\omega$  in entire *Drosophila* genome. **C** Violin plot representation of distribution of  $\omega$  value for non-T2D (red) and T2D (cyan) genes. Black dot in the center of the

red and cyan violin plot represents the mean  $\omega$  value. **D** Pairwise assessments of dS and dN substitution rates in T2D genes. **E** Saturation plot in T2D genes. **F**  $\omega$  value of T2D genes were plotted against sequence divergence ( $t$ )

## Gene ontology and pathway enrichment analysis

STRING database (Szklarczyk et al. 2017) was utilized for detecting gene ontology (cellular component, biological process, molecular function) and pathway enrichment analysis associated with key genes. Result having gene count > 2 and FDR < 0.05 were considered significant.

## Generation of the three-dimensional structure of proteins

The coding sequence of each key genes was submitted to the EMBOSS Transeq tool ([https://www.ebi.ac.uk/Tools/st/emboss\\_transeq/](https://www.ebi.ac.uk/Tools/st/emboss_transeq/)) for translating nucleic acid sequences to their corresponding peptide sequences. The protein sequence of each key candidate genes was subject to NCBI's protein BLAST (Altschul et al. 1990), separately, for detecting their nearby homologous structure against Protein Data Bank (PDB). Depending on maximum sequence identity, query coverage, or lower e-value, if their homologous structure

were available in Protein Data Bank, they were retrieved manually. If homologous structures are absent in Protein Data Bank, each protein sequence was submitted separately to the GalaxyWEB server (<https://galaxy.seoklab.org/>) for building their three-dimensional model. Ramachandran plot produced via PROCHECK (Laskowski et al. 1993) and Z scores computed through the ProSA-web tool (Wiederstein and Sippl 2007) were employed to validate the geometry of modeled protein structure.

## Molecular dynamics (MD) simulations of proteins

For understanding structural characteristics of the protein encoded via each key genes distinctly, molecular dynamics simulations of each protein for 200 ns were done via Gromos96-43a1 force field of “GROningen Machine for Chemical Simulations” (GROMACS 5.1) (Abraham et al. 2015), individually. If proteins encoded through key candidate genes are situated in the cytoplasm, cubic boxes bearing SPC216 water molecules employed for solvating

protein (Gupta and Vadde 2018; Gouda et al. 2019). If proteins produced via key candidate genes are situated in the cell membrane, three-dimension structure of each protein was implanted into equilibrated bilayer of “dipalmitoyl phosphatidylcholine” employing “g\_membed” tool (Wolf et al. 2010) of GROMACS utilizing parameters for “Berger lipids” generated from Berger, Edholm, and Jahnig (Berger et al. 1997). Further, the solvation of the membrane systems was carried out by creating a local copy of vdwradii.dat and modifying C value to 0.375 from 0.15 (Gupta and Vadde 2019b; Gupta et al. 2019). By making these changes, solvate assign carbon atoms sufficiently large van der Waals radius, which in turn makes water augmentation within the lipids less likely (Lemkul 2015). Further, neutralization of the entire system was performed through adding suitable ions employing genion application of GROMACS.

To discard faulty steric conflicts and van der Waals contacts protein, energy-minimization performed at an initial stage through steepest descent of 3,000 steps with 0.01 nm energy step size. The energy minimisation step was designed to halt when the maximum force reaches less than 1000 kJ/mol/nm. To equilibrate the complete system, solute was subjected to constant “number of particles, volume and temperature” (NVT) conditions for 100 ps at 300 K, subsequently followed through 100 ps under constant “number of particles, pressure, and temperature” (NPT) conditions up to 200 ns. All covalent bonds were moderated through the “Linear Constraint Solver” (LINCS) algorithm. Last step of molecular dynamics of the electrostatic interactions was computed through “Particle Mesh Ewald” (PME) method. For each and every step, dynamics simulation was allotted 100,000,000 steps with an energy step size of 0.02 fs (200 ns). The protein atoms were harmonically constrained during solvent equilibration (Donde et al. 2019). Final MD trajectories, along with the quality of simulations, were estimated through GROMACS5.1. Xmgrace (Turner 2005) program was employed for generating two-dimensional plots and trajectory analysis.

## Result

### File preprocessing and orthologs search

Initial inspection of each aligned CDS file reveals that there are more than ~7304 genes across the phylogeny, but 7304 orthologs are present across all groups in the *Drosophila* genus. However, out of 7304, only 5679 genes are present in all 12 species of *Drosophila*. Further, investigation via the DIOPT-DIST tool suggested that, out of 5679, only 202 orthologs of human protein-coding T2D genes are present in *Drosophila* (Supplementary File 1). Thus, aligned coding

sequence files of only 5679 (202 T2D and 5476 non-T2D) genes were considered for downstream analysis.

### Identification of selection pressure acting on T2D genes and its sites in the *Drosophila* genus

The result obtained from the M0 model suggests that the  $\omega$  value of all genes is less than 1 (Table 1 and Fig. 1B), which in turn supports that almost the entire genome of the *Drosophila* genus is evolving under purifying selection. However, the mean  $\omega$  of T2D genes (0.063) is slightly less than non-T2D genes (0.064) (Fig. 1C). In T2D genes,  $dS$  (range 0.032–25.827) is greater than  $dN$  values (range 0.000–1.431) (Fig. 1D). Saturation plot of T2D genes reveals strong saturation of both transitional as well as transversal substitutions till 0.05 ( $\leq 1$ ) (Fig. 1E); thereby suggesting the presence of multiple substitutions as well as plausible homoplasy (Farfán et al. 2009). Figure 1F depicts a negative relationship between  $\omega$  & sequence divergence. Further quantile regression analysis between  $\omega$  values of T2D and non-T2D genes suggests that T2D is evolving under strong purifying selection in *Drosophila* genus ( $p$ -value = 0.044) (Table 1).

It is pertinent to note that  $dS$  across 12 species of is higher than 1 in almost all genes. Hence, to avoid misestimation of  $\omega$  in each species of *Drosophila*, at first, we split our datasets into four groups, namely Group A (*D. melanogaster*, *D. sechellia* and *D. simulans*), Group B (*D. erecta*, *D. yakuba*), Group C (*D. persimilis* and *D. pseudoobscura*) and Group D (*D. virilis*, *D. grimshawi* and *D. mojavensis*). Subsequently,  $dS$  and  $dN$  value of each gene of every group was estimated using the M0 model. Result obtained from “M0

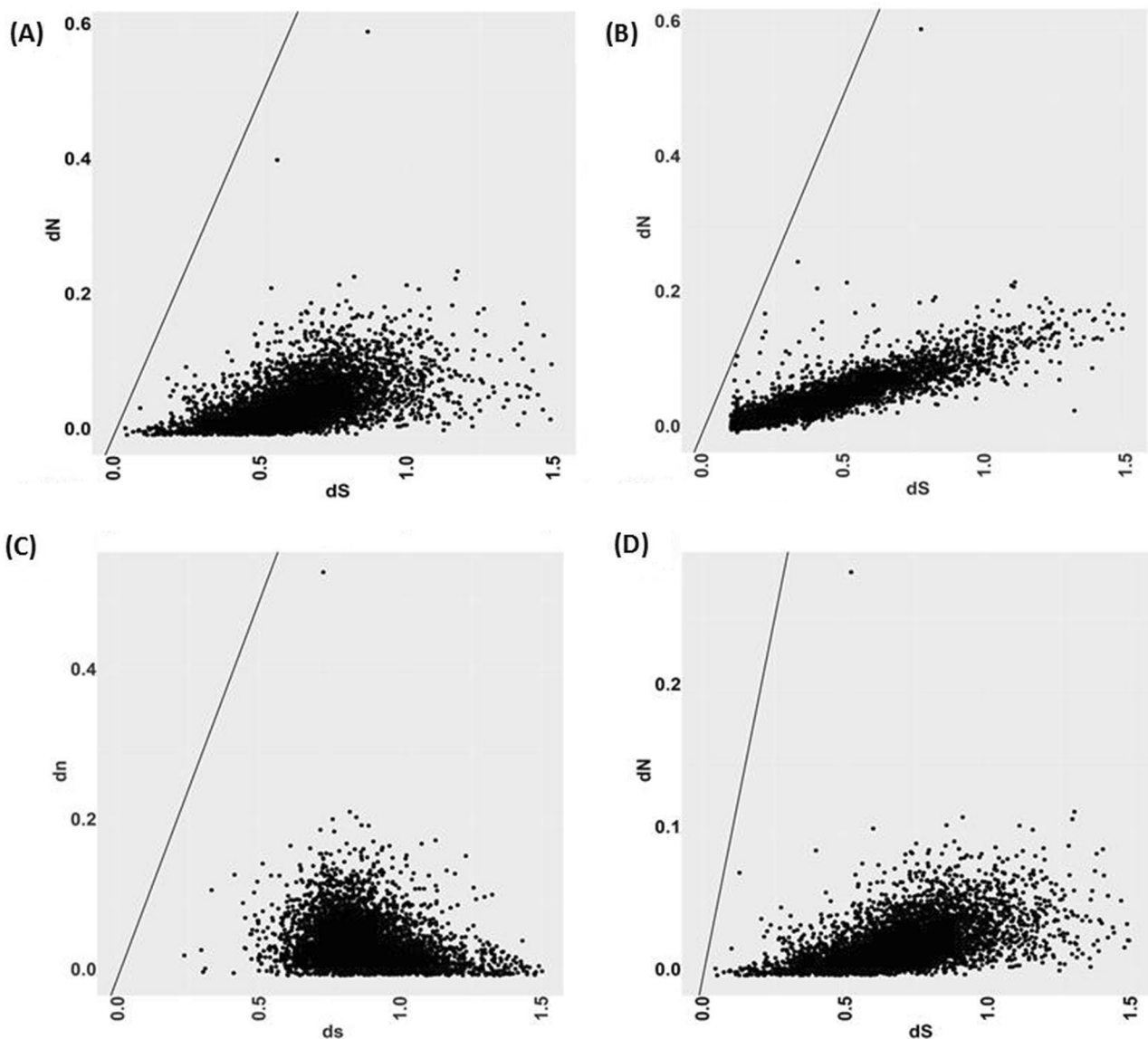
**Table 1** Result obtained from quantile regression analysis between  $\omega$  values of T2D and non-T2D genes in different species of *Drosophila*

Organism	Value	Std. error	t value	Pr(> t )
<i>Drosophila</i> genus	− 0.005	0.003	− 2.015	0.044
Group A				
<i>D. melanogaster</i>	− 0.008	0.002	− 3.137	0.002
<i>D. simulans</i>	− 0.017	0.008	− 2.251	0.024
<i>D. sechellia</i>	− 0.021	0.007	− 2.911	0.004
Group B				
<i>D. yakuba</i>	− 0.005	0.004	− 1.276	0.202
<i>D. erecta</i>	− 0.009	0.010	− 0.850	0.396
Group C				
<i>D. pseudoobscura</i>	− 0.005	0.005	− 1.130	0.258
<i>D. persimilis</i>	− 0.030	0.036	− 0.839	0.402
Group D				
<i>D. virilis</i>	− 0.005	0.003	− 1.593	0.111
<i>D. mojavensis</i>	− 0.007	0.005	− 1.303	0.193
<i>D. grimshawi</i>	− 0.007	0.005	− 1.248	0.212

models” reveals that dS value in each range between 0 and 1.5 (Fig. 2). Later, genes with  $0.05 < dS < 1$  was only considered for "Branch-site models" analysis to get a better estimation of  $\omega$  in each group distinctly. It is pertinent to note that dS value of all 202 T2D genes in all four group range between 0.05 and 1.

The result obtained from "Branch-site models" reveals that mean  $\omega$  of T2D genes is less than non-T2D genes in all species of each four group (Supplementary File II and Table 1). However, quantile regression analysis between  $\omega$  of T2D and non-T2D genes in each species separately suggests that T2D is evolving significantly under strong

purifying selection only in GroupA (p-value  $< 0.05$ ) (Table 1). Hence, *D. melanogaster*, *D. sechellia*, and *D. simulans* serve as a better model to study T2D in comparison to others. Later, LRT between Model 8 and 7 in all four groups suggests that few sites only in three T2D genes, namely, *CG8051* (ASN7, ALA71, THR323 & HIS330), *ZnT35C* (VAL22, SER174, ALA177, THR227) and *kar* (ASN496 & ALA499), of GroupA experience positive selection (Supplementary file III). No significant result found in GroupB, GroupC and GroupD. Thus, these three genes are key genes and may play an important role in maintaining normal insulin level in the body.



**Fig. 2** Pairwise assessments of dS and dN substitution rates in T2D genes of **A** Group A (*D. melanogaster*, *D. sechellia* and *D. simulans*), **B** Group B (*D. erecta*, *D. yakuba*), **C** Group C (*D. persimilis* and *D. pseudoobscura*) and **D** Group D (*D. virilis*, *D. grimshawi* and *D. mojavensis*)

## Gene ontology and pathway enrichment analysis

Analysis of three key genes via the STRING database reveals that the main molecular function associated with them are monocarboxylic acid transmembrane transporter activity, ion transmembrane transporter activity and inorganic molecular entity transmembrane transporter activity (Fig. 3A). All three key genes mainly reside in the integral component of the membrane. The main biological process associated with them is monocarboxylic acid transport, ion transmembrane transport, and carboxylic acid transmembrane transport. The main pathway associated with these key genes is SLC-mediated transmembrane transport.

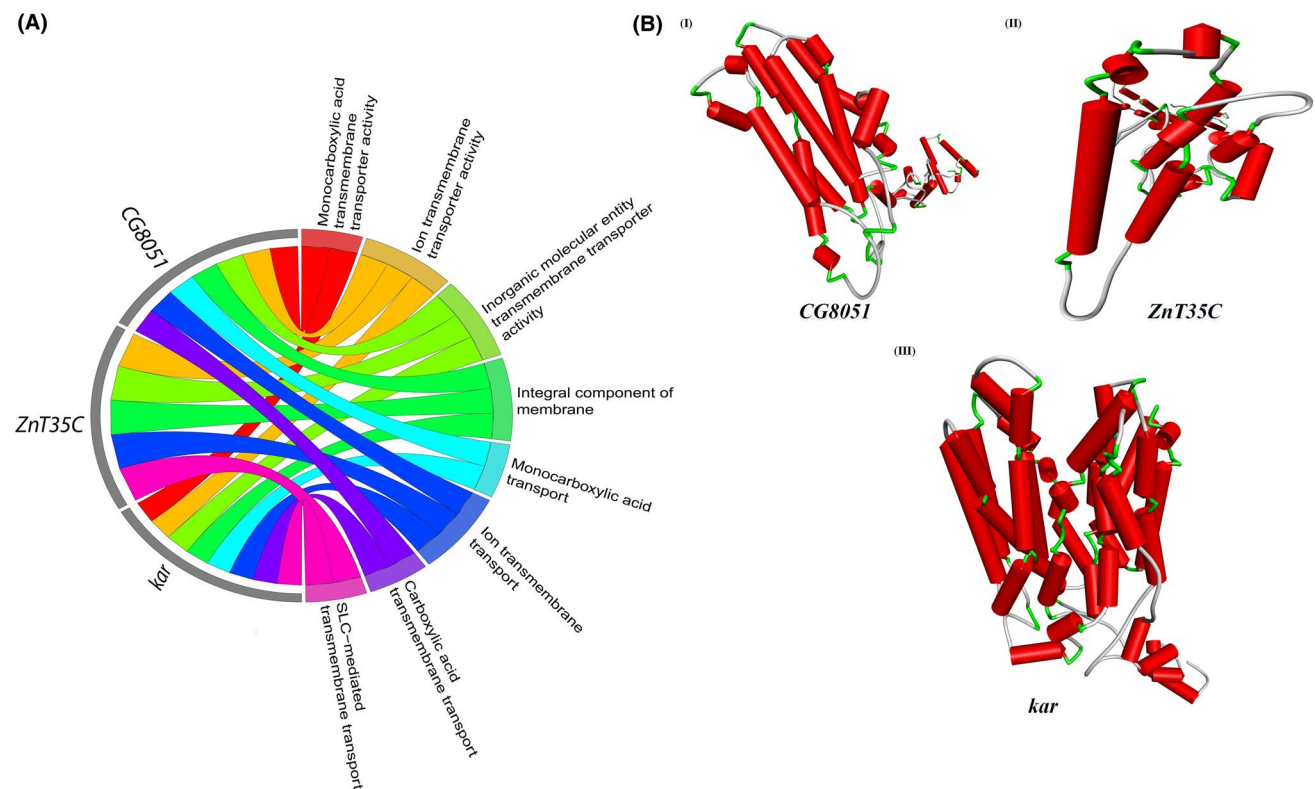
## Generation of the three-dimensional structure of proteins

As no homologous structure of proteins encoded via these three key genes was identified in Protein Data Bank, the protein sequence of each key gene was submitted to the GalaxyWEB server separately. Generated structures of *CG805*, *ZnT35C*, and *kar* are made up of only 24, 19, and 36  $\alpha$ -helices and loops, respectively (Fig. 3B, I–III). Further, validation of the three-dimensional structure of each protein via PROCHEK suggests that overall, 99.762%, 98.000% and 99.000%, of *CG8051*, *ZnT35C*, and *kar* residues,

respectively, are present in allowed regions (Fig. 4). Z scores of *CG8051* (1.800), *ZnT35C* (0.560) and *kar* (1.540) also range between -10 and 10, thereby proposing that stereochemical geometry of the generated models is sensibly good.

## MD simulations of proteins

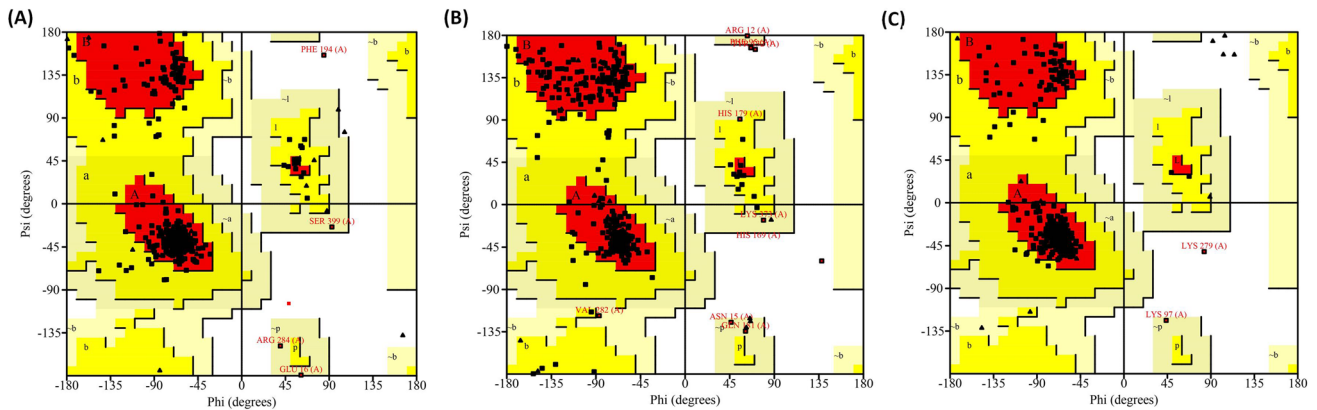
To understand structural characteristics, MD trajectories of three-dimensional structure of all three proteins for 200 ns were performed separately. During energy minimization, the final potential energy of *CG8051*, *ZnT35C*, and *kar*, was -2,311,545.250 kJ/mol, -4,860,415.506 kJ/mol, and -3,469,797.000 kJ/mol, respectively. The temperature of all three proteins varies between 298 and 301 K with mean 299.999 K. Mean pressure of *CG8051*, *ZnT35C*, and *kar*, is -0.175 bar, 1.380 bar, and 0.962 bar, respectively. Mean density of *CG8051*, *ZnT35C*, and *kar* is 978.826, 1014.623, and 998.654, respectively. In *CG8051*, Rg varies between 10.042 Å and 10.132 Å with a mean Rg of 10.09 Å. In *ZnT35C*, Rg varies between 5.657 Å and 5.674 Å with mean Rg of 5.665 Å. In *kar*, Rg varies between 7.362 Å and 7.497 Å with a mean Rg of 7.429 Å. Mean RMSD and RMSF of *CG8051* < *kar* < *ZnT35C*. However, RMSF of amino acids towards the N-terminal is constrained. Amino acid experiencing the highest fluctuation during simulation in *CG8051* are PRO257, ARG273, THR323, HIS330,



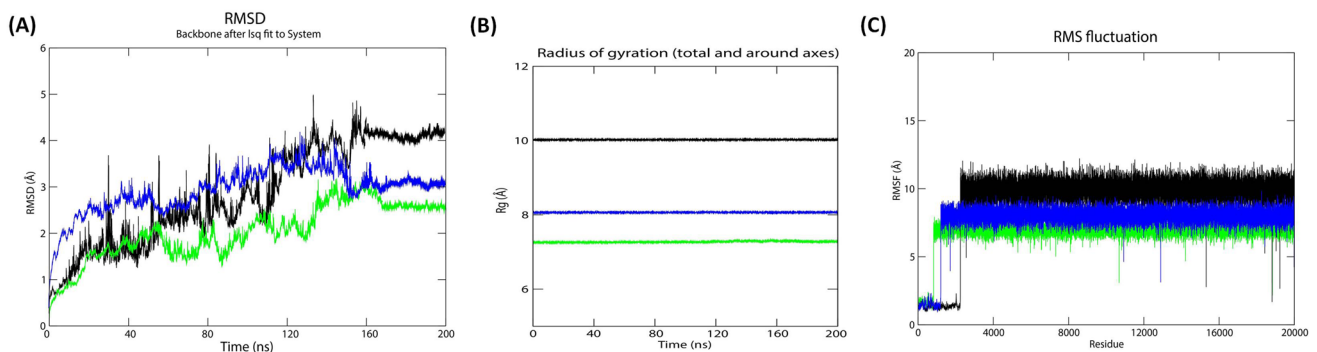
**Fig. 3** **A** Ontology and **B** three-dimensional structure of proteins encoded by each of the three key genes

GLU337, and THR372. Amino acid experiencing the highest fluctuation during simulation in *ZnT35* are VAL266, THR277, CY269, and LEU364. Amino acid experiencing

the highest fluctuation during simulation in *kar* are ALA236, AGR330, THR359, and ALA499 (Fig. 5). 'Cross-correlation matrix' of the C- $\alpha$  displacement (Fig. 6) and the 'free energy

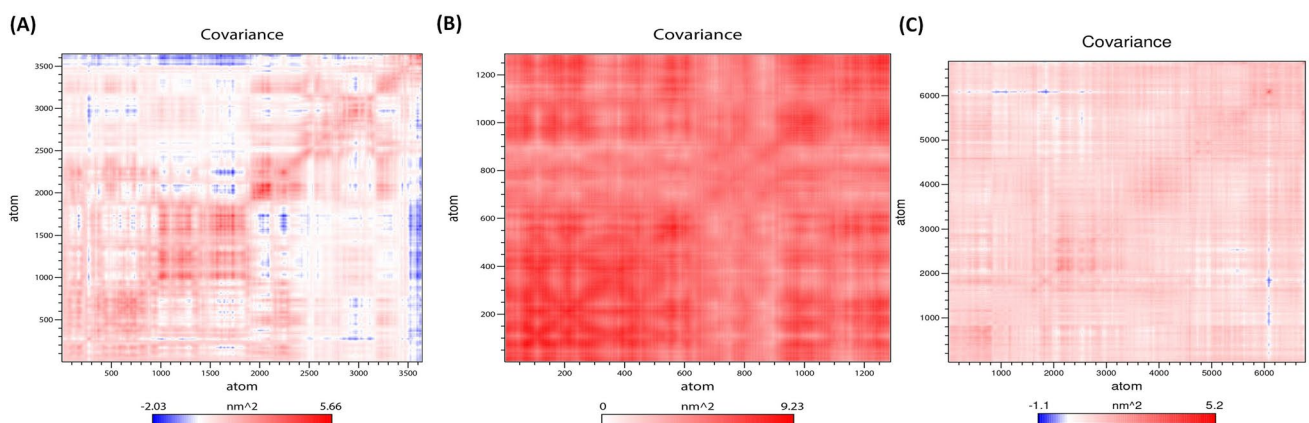


**Fig. 4** Ramachandran plot of modeled proteins encoded via **A** *CG8051*, **B** *ZnT35C*, and **C** *kar* generated via PROCHECK



**Fig. 5** The stability parameters for key T2D protein during 200 ns: **A** RMSD of C- $\alpha$  **B** RMSF of C- $\alpha$  and **C** Radius of gyration of C- $\alpha$ . The trajectory projected to the two-dimensional space. Black, light green,

and blue, lines represent proteins encoded via *CG8051*, *ZnT35C*, and *kar* during 200 ns, respectively



**Fig. 6** Comparative study of cross-correlation matrices of C- $\alpha$  atoms of modeled proteins encoded via **A** *CG8051*, **B** *ZnT35C*, and **C** *kar* during 200 ns simulation. The range of motion indicated by various

colors in the panel. Red indicates a positive correlation, whereas blue indicates anti-correlation



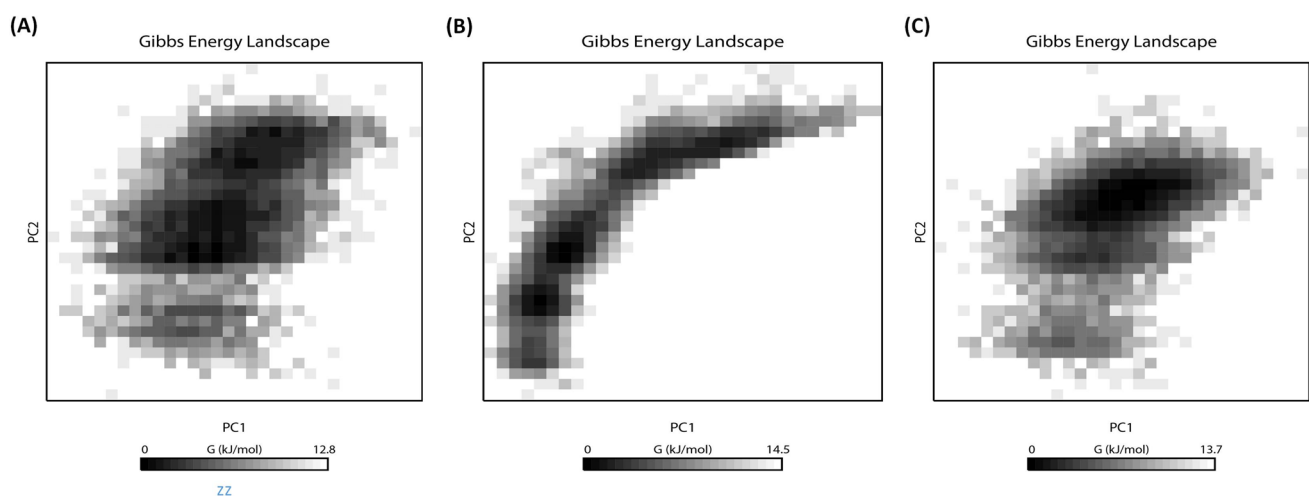
landscape’ analysis (Fig. 7) revealed that all residues present within *CG8051* experience random movement while *ZnT35* experiences constrain movements under wild condition.

## Discussion

As both incident of T2D, as well as mortal rate due to T2D, is increasing dramatically every year and imposing huge financial burden on almost every country, there is always an urge to look for new approaches or technology which may enable us to detect key genes and pathway that play key role in the T2D development. For instance, Hu and team (2009) performed GWAS analysis and detected SNPs in *PPARG* (rs1801282), *KCNJ11* (rs5219), *CDKAL1* (rs10946398, rs7754840, rs9460546, rs7756992 & rs9465871), *CDKN2A–CDKN2B* (rs564398 & rs10811161), *IDE-KIF11-HHEX* (rs10509645, rs1111875 and rs10748582), *IGF2BP2* (rs7651090) and *SLC30A8* (rs13266634) that are responsible for causing T2D (Hu et al. 2009). Additionally, several authors are also employing evolutionary approaches for unmasking the pathophysiology and molecular mechanism associated with T2D in a more comprehensive way. For instance, in 2017, Little and team tested the hypothesis that “natural selection is associated with type 2 diabetes (T2D)-associated mortality and fertility in a rural, isolated Zapotec community in the Valley of Oaxaca, southern Mexico” and reported that frequency of T2D-mortality increases with decrease in natural selection as well as favoured offspring survival of non-T2D descendants (Little et al. 2017). Hence, evolutionary comparative sequence analysis is a powerful way of unraveling the mechanisms that shaped contemporary genetic diversity.

Biochemical pathways involved in growth and metabolism are ancient and well conserved across the animal kingdom. Due to conservation between humans & other organisms at both molecular as well as physiological levels, these organisms may be utilized for understanding the real mechanism associated with T2D development in humans. Numerous T2D associated studies have also been performed in various model organisms, like KK mice and *Drosophila melanogaster* (King 2012; Murillo-Maldonado and Riesgo-Escovar 2017). Most of the animal models of T2D are obese, mimicking human conditions where obesity is the main cause for developing T2D (King 2012; Murillo-Maldonado and Riesgo-Escovar 2017). The *fa/fa* rats and *ob/ob* mice are one of the best examples for the same. Other model organisms, for instance, *Psammomys obesus* (the Israeli sand rat) and *db/db* mouse, develop hyperglycemia rapidly because their  $\beta$ -cells are incapable of maintaining a high concentration of insulin secretion required throughout life. The study of these animal models may provide significant insight why few humans with severe obesity never develop T2D, while others are more risk at developing hyperglycemic despite modest insulin resistance and obesity (Rees and Alcolado 2005). The zebrafish model also showed a better response to the anti-diabetic drug, namely metformin, and glibenclamide, proposing that zebrafish can also be utilized as a model organism towards understanding the mechanism of T2D in human. However, the organisms, especially mouse, rat, and dog, have strict ethical guidelines for carrying out research (Rees and Alcolado 2005). Additionally, life span of these organisms are also large, and hence, special care is also required for maintaining these organisms.

Owing to smaller genome size and short life span, *Drosophila melanogaster* serves as one of the best models



**Fig. 7** Projections of the free energy landscape of protein encoded via **A** *CG8051*, **B** *ZnT35C*, and **C** *kar* during 200 ns simulation. Various colors in the panel indicate the range of motion, where dark black

indicates the lowest energy configuration, and white shows the highest energy configuration

for studying any human diseases. The main advantage of employing *Drosophila melanogaster* is that, unlike other organisms (e.g., mouse and dog), there are either no or few ethical issues surrounding their use (Jennings 2011). Insulin producing cells (IPCs) of *Drosophila* is equivalent to mammalian Langerhans' islets  $\beta$  pancreatic cells (Alfa and Kim 2016; Graham and Pick 2017). However, unlike vertebrates, which have one insulin gene, *Drosophila* genome encodes seven different insulin-like peptides (ILPs) (Álvarez-Rendón et al. 2018). The ILP2 peptide has the highest homology with insulin gene of the vertebrate and is produced along with ILP1, ILP3 & ILP5 in the IPCs located in the brain (Álvarez-Rendón et al. 2018). ILP2 is also expressed in the imaginal discs and the salivary glands. ILP4, ILP5, and ILP6 are expressed in the midgut, and ILP7 is expressed in the ventral nerve chord. These seven ILPs function together with *Drosophila's* insulin-like receptor (InR) trigger a cascade of intracellular events facilitated via conserved apparatuses of the insulin/IGF pathway comprising of the insulin receptor substrate (IRS) Chico, the insulin signaling antagonist PTEN, PKB/Akt kinase, PI3K, and dFOXO (the single FOXO orthologs) (Baker and Thummel 2007).

In a normal feeding environment, three *ILP* genes, namely, ILP2, ILP3 and ILP5, are expressed within median neurosecretory cells of the brain and regulate sugar level in circulating blood. In response to reduced dietary carbohydrate concentration, the expression of ILP3 and ILP5 decreases in IPCs, suggesting that ILP concentration can respond to particular nutritional indications like insulin in humans (Baker and Thummel 2007). Additionally, some studies reported that removal of the insulin-producing cells causes hyperglycemia (Grönke et al. 2010). Like glucagon secretion from pancreatic  $\alpha$ -cells in mammals, insect adipokinetic hormone (AKH) counterbalances the actions of insulin via triggering glycogen phosphorylase, enhancing circulating sugars and decreasing fat body glycogen. *Akh* is expressed in the main endocrine organ of the insects, namely the corpora cardiaca region of the ring gland, which is in direct contacts with the hearts and IPCs. Removal of corpora cardiaca specific cell removes *Akh* function, which in turn reduces the concentration of circulating trehalose but has no significant effect on the stored concentration of lipid or glucose (Lee and Park 2004; Baker and Thummel 2007). However, ectopic expression of *Akh* in the fat body, the primary target tissue of *Akh*, causes hypertrehalosemia and lipolysis, which in turn reduces the amount of stored lipid (Lee and Park 2004). All these earlier studies reported that the central regulatory functions of insulin and glucagon are conserved throughout evolution and supported that *Drosophila* can be utilized as a valid model organisms for functional studies of glucose homeostasis as well as the underlying mechanisms modulating the onset of diabetes. Thus, in present study, authors made an attempt to re-analyzed the publicly

available T2D gene sequences of *Drosophila* for studying evolutionary processes responsible for shaping genetic make-up of T2D genes in genus *Drosophila*.

Result obtained reveals that there are only 202 orthologs of human protein-coding T2D genes in *Drosophila* genus. Few human T2D genes like *ARF5*, *LIPC*, *CPA6*, *CCNQ*, *KCNJ11*, and *GALNT14* have more than one orthologs in *Drosophila* (Supplementary File I). This might be because *Drosophila* may have underwent an additional round of whole-genome duplication during evolution (Maurer et al. 2015). Further analysis via M0 model of CODEML reveals that all T2D genes present in the *Drosophila* genus are evolving significantly under purifying selection ( $p$ -value < 0.05). Earlier studies have also reported that in comparison to younger proteins, ancient proteins exhibit stronger purifying selection; thereby indicating T2D genes is ancient (Domazet-Loso and Tautz 2003, 2008). The functions of ancient genes, like T2D genes, are highly optimized as well as conserved, and they are likely to have already exhausted all beneficial mutations in recent times. Thus they are expected to evolve under purifying selection and fix only neutral and/or nearly neutral mutations (Vishnoi et al. 2010). Our results is also in accordance with Blekhman and the team, who also demonstrated that genes associated with Mendelian and complex diseases are under purifying selection (Blekhman et al. 2008).

Further, since  $dS$  across 12 species of is higher than 1 in almost all genes, dataset was divided into four groups and  $\omega$  of each species in each group was estimated separately using "Branch-site models". Result obtained revealed that T2D is evolving significantly under strong purifying selection only in GroupA ( $p$ -value < 0.05). Hence, *D. melanogaster*, *D. sechellia* and *D. simulans* serve as a better model to study T2D in comparison to other (Table 1). This result is in accordance with earlier studies where authors have reported that evolution shape each branch of the phylogeny distinctly because the number of rates of nonsynonymous and synonymous substitutions varies across a sequence and species (Wong et al. 2008). LRT between Model 8 and 7 in all the four group, separately, suggests that few sites only in three key T2D genes, namely, *CG8051*, *ZnT35C*, and *kar*, of GroupA experience positive selection (Supplementary file II). This is in accordance with earlier studies where authors have reported that few sites in genes that are evolving under purifying selection may also experience adaptive change occasionally (Yang and Bielawski 2000). Earlier studies have also reported that positive selection is the preservation and spread of beneficial mutations throughout the population. Identifying positively selected protein or its site in any branch of a phylogeny suggests that there is a selective advantage of positively selected protein or its site over another branch of a phylogeny. This selective advantage may be in response to change in the various external and

internal phenomena, for instance, diet, disease, and adaptation to several ecological niches (Morgan et al. 2012). For instance, the genetic adaptations to the low-salt environment in ancestral populations is a threat to hypertension in present populations residing in a high-salt environment (Balaresque et al. 2007). This salt retention adaptive trait enables ancient humans, consuming low levels of dietary salt to survive in hot and humid areas (Balaresque et al. 2007). Earlier studies have reported that numerous olfactory system genes are positively selected in organisms if odor and pheromone perception is crucial for its reproduction as well as survival (Ngai et al. 1993; Willett 2000; Emes et al. 2004; Krieger and Ross 2002). Such sites point to functionally important gene's regions and, hence, are of potential interest to protein engineers who alter proteins to produce new functions. Information about positively selected protein and their sites are highly required for our understanding of functionally significant amino acids in any protein sequence and their role in protein functional shift (Yang and Bielawski 2000).

Gene ontology and pathway enrichment analysis reveals that these three key genes encode membranes proteins and are mainly involved in the ion transport (Fig. 3A). Earlier studies have also reported that ion channels as well as transporters proteins play key roles in both excitable cells, e.g., skeletal, cardiac, neurons, as well as endocrine cells, and non-excitable cells, e.g., liver (Spires et al. 2019). In human pancreatic  $\beta$ -cells,  $K_{ATP}$  channels modulate the membrane potential of the  $\beta$ -cell membrane, which in turn regulates insulin secretion (Spires et al. 2019; Gupta and Vadde 2020a). For instance, several other studies have reported that in humans, ZnT8 transporter protein resides on dense-core vesicles in pancreatic  $\beta$  cells and loads  $Zn^{2+}$  into these secretory compartments, where it binds with and stabilizes a hexameric form of insulin (O'Halloran et al. 2013). These ZnT8 transporters are mainly responsible for the efflux of zinc from the cytosol to intracellular vesicles, unlike the functions of zinc importers (ZiPs; *SLC39*), which are responsible for zinc influx into the cytosol as well as zinc-binding proteins, like metallothionein. This co-ordination between the function of zinc transporter and importers maintain zinc level in cytosol (Rutter and Chimienti 2015; Gupta and Vadde 2020b). However, mutations in this two protein, namely  $K_{ATP}$  channels and ZnT8 transporter, is reported to disrupt their normal functions, which in turn cause T2D (Gupta and Vadde, 2020a; b).

Further, molecular dynamics studies suggest that all these three key genes are mainly comprised of  $\alpha$ -helices and loops (Fig. 3B). Out of three, *CG8051* experiences more random movement while ZnT35C experience constrains movement under normal conditions. This might be due to less potential energy and higher pressure in ZnT35C. Movement of N-terminal residues of all three key genes is more constrained, thereby supporting that the N-terminal region of all these

three proteins is insignificant during protein–ligand interaction. This finding is in accordance with our earlier studies in the human ortholog of *ZnT35C*, i.e. zinc transporter ZnT8. Movement of ZnT8 protein was also found to be constrained under normal conditions (Gupta and Vadde 2020b). N-terminal region of ZnT8 was also found to be insignificant during protein–ligand interaction (Gupta and Vadde 2020b). RSMF analysis reveals that, out of all positively selected sites, only ARG273 and THR323 in *CG805*, THR277 in *ZnT35C*, ALA499 in *kar* experiences highest fluctuations; thereby supporting their importance during protein–ligand interaction. This, in turn, helps to modulate the normal metabolic function of the body. Thus, in summary, as T2D disease is ancient, they are evolving under purifying selection in the *Drosophila* genus. Hence, the function of T2D genes is highly conserved throughout evolution. However, few sites in membrane proteins encoded T2D genes, like *CG8051*, *ZnT35C*, and *kar*, are still evolving under positive selection in few species of *Drosophila*, like, *D. melanogaster*, *D. sechellia* and *D. simulans*; this might be due to adaptive (positive) evolution in response to changes in various external mechanisms, for instance, response to disease & adaptation to several ecological niches, and internal mechanisms (compensatory mutations and co-evolution).

## Conclusions

In conclusion, as T2D genes are ancient, they are evolving under purifying selection. Hence, there is almost no or very little scope for new nonsynonymous mutations in T2D genes, and the functions of T2D genes are highly conserved throughout evolution. However, few sites in membrane proteins encoded via few T2D genes, like *CG8051*, *ZnT35C*, and *kar*, are still evolving under positive selection in certain scenarios, which might be due to adaptive (positive) evolution in response to changes in various external mechanisms, for instance, response to disease and adaptation to several ecological niches, and internal mechanisms (compensatory mutations and co-evolution). This study provides a new perspective on an understanding of the evolution of the T2D gene. In the near future, information obtained from the present study will be highly useful in the field of evolutionary medicine, as well as in the drug discovery process.

**Acknowledgements** Authors thank Dr. Julien Y. Dutheil, Group Leader, Max Planck Institute for Evolutionary Biology, Plön, Germany, and Mr. Jan Benzenberg, Incident Manager, Carrier Radio Access Networks (Germany), for assistance with the computational facility. Author also would like to thanks Mrs. G Radhika Prashanna, English Lecturer, Government Polytechnique College Anantpur, India for proofreading the manuscript.

## Compliance with ethical standards

**Conflict of interest** The authors of this manuscript declare no conflict of interest.

**Ethical approval** This article is a computational work and does not contain any studies with human participants or animals.

**Informed consent** This article is computational work. Hence, this article does not require consent from any individual or organization.

## References

- Abraham MJ, Murtola T, Schulz R et al (2015) GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1–2:19–25. <https://doi.org/10.1016/j.softx.2015.06.001>
- Ahmed-Braimah YH, Unckless RL, Clark AG (2017) Evolutionary dynamics of male reproductive genes in the drosophila virilis subgroup. *G3* 7:3145–3155. <https://doi.org/10.1534/g3.117.1136>
- Al-Daghri NM, Pontremoli C, Cagliani R et al (2017) Susceptibility to type 2 diabetes may be modulated by haplotypes in G6PC2, a target of positive selection. *BMC Evol Biol*. <https://doi.org/10.1186/s12862-017-0897-z>
- Alfa RW, Kim SK (2016) Using *Drosophila* to discover mechanisms underlying type 2 diabetes. *Dis Model Mech* 9:365–376. <https://doi.org/10.1242/dmm.023887>
- Altschul SF, Gish W, Miller W et al (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- Álvarez-Rendón JP, Salceda R, Riesgo-Escovar JR (2018) *Drosophila melanogaster* as a model for diabetes type 2 progression. In: *BioMed research international*. <https://www.hindawi.com/journals/bmri/2018/1417528/>. Accessed 2 Dec 2018
- Baker KD, Thummel CS (2007) Diabetic larvae and obese flies: emerging studies of metabolism in *Drosophila*. *Cell Metab* 6:257–266. <https://doi.org/10.1016/j.cmet.2007.09.002>
- Balaresque PL, Ballereau SJ, Jobling MA (2007) Challenges in human genetic diversity: demographic history and adaptation. *Hum Mol Genet* 16(2):R134–139. <https://doi.org/10.1093/hmg/ddm242>
- Baynest HW (2015) Classification, pathophysiology, diagnosis and management of diabetes mellitus. *J Diabetes Metab*. <https://doi.org/10.4172/2155-6156.1000541>
- Berger O, Edholm O, Jähnig F (1997) Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature. *Biophys J* 72:2002–2013. [https://doi.org/10.1016/S0006-3495\(97\)78845-3](https://doi.org/10.1016/S0006-3495(97)78845-3)
- Blekhman R, Man O, Herrmann L et al (2008) Natural selection on genes that underlie human disease susceptibility. *Curr Biol* 18:883–889. <https://doi.org/10.1016/j.cub.2008.04.074>
- Domazet-Loso T, Tautz D (2003) An evolutionary analysis of orphan genes in *Drosophila*. *Genome Res* 13:2213–2219. <https://doi.org/10.1101/gr.1311003>
- Domazet-Loso T, Tautz D (2008) An ancient evolutionary origin of genes associated with human genetic diseases. *Mol Biol Evol* 25:2699–2707. <https://doi.org/10.1093/molbev/msn214>
- Donde R, Gupta MK, Gouda G et al (2019) Computational characterization of structural and functional roles of DREB1A, DREB1B and DREB1C in enhancing cold tolerance in rice plant. *Amino Acids* 51:839–853. <https://doi.org/10.1007/s00726-019-02727-0>
- Drosophila* 12 Genomes Consortium (2007) Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450:203–218. <https://doi.org/10.1038/nature06341>
- Emes RD, Beatson SA, Ponting CP, Goodstadt L (2004) Evolution and comparative genomics of odorant- and pheromone-associated genes in rodents. *Genome Res* 14:591–602. <https://doi.org/10.1101/gr.1940604>
- Farfán M, Miñana-Galbis D, Fusté MC, Lorén JG (2009) Divergent evolution and purifying selection of the *flaA* gene sequences in *Aeromonas*. *Biol Direct* 4:23. <https://doi.org/10.1186/1745-6150-4-23>
- Fischman BJ, Woodard SH, Robinson GE (2011) Molecular evolutionary analyses of insect societies. *PNAS* 108:10847–10854. <https://doi.org/10.1073/pnas.1100301108>
- Anastasia G, Butlin Roger K, Jordan William C, Ritchie Michael G (2009) Sites of evolutionary divergence differ between olfactory and gustatory receptors of *Drosophila*. *Biol Lett* 5:244–247. <https://doi.org/10.1098/rsbl.2008.0723>
- Gouda G, Gupta MK, Donde R et al (2019) Computational approach towards understanding structural and functional role of cytokinin oxidase/dehydrogenase 2 (CKX2) in enhancing grain yield in rice plant. *J Biomol Struct Dyn*. <https://doi.org/10.1080/07391102.2019.1597771>
- Graham P, Pick L (2017) *Drosophila* as a model for diabetes and diseases of insulin resistance. *Curr Top Dev Biol* 121:397–419. <https://doi.org/10.1016/bs.ctdb.2016.07.011>
- Grönke S, Clarke D-F, Broughton S et al (2010) Molecular evolution and functional characterization of drosophila insulin-like peptides. *PLoS Genet* 6:e1000857. <https://doi.org/10.1371/journal.pgen.1000857>
- Grunspan DZ, Nesse RM, Barnes ME, Brownell SE (2017) Core principles of evolutionary medicine. *Evol Med Public Health* 2018:13–23. <https://doi.org/10.1093/emph/eox025>
- Gupta MK, Vadde R (2018) In silico identification of natural product inhibitors for  $\gamma$ -secretase activating protein, a therapeutic target for Alzheimer's disease. *J Cell Biochem*. <https://doi.org/10.1002/jcb.28316>
- Gupta MK, Vadde R (2019a) Genetic basis of adaptation and maladaptation via balancing selection. *Zoology*. <https://doi.org/10.1016/j.zool.2019.125693>
- Gupta MK, Vadde R (2019b) Insights into the structure-function relationship of both wild and mutant Zinc transporter ZnT8 in human: a computational structural biology approach. *J Biomol Struct Dyn*. <https://doi.org/10.1080/07391102.2019.1567391>
- Gupta MK, Vadde R (2020a) A computational structural biology study to understand the impact of mutation on structure–function relationship of inward-rectifier potassium ion channel Kir6.2 in human. *J Biomol Struct Dyn*. <https://doi.org/10.1080/07391102.2020.1733666>
- Gupta MK, Vadde R (2020b) Insights into the structure–function relationship of both wild and mutant zinc transporter ZnT8 in human: a computational structural biology approach. *J Biomol Struct Dyn* 38:137–151. <https://doi.org/10.1080/07391102.2019.1567391>
- Gupta MK, Vadde R, Gouda G et al (2019) Computational approach to understand molecular mechanism involved in BPH resistance in Bt- rice plant. *J Mol Graph Model* 88:209–220. <https://doi.org/10.1016/j.jmgm.2019.01.018>
- Hill T, Koseva BS, Unckless RL (2019) The genome of *Drosophila innubila* reveals lineage-specific patterns of selection in immune genes. *Mol Biol Evol*. <https://doi.org/10.1093/molbev/msz059>
- Hu C, Zhang R, Wang C et al (2009) PPARG, KCNJ11, CDKAL1, CDKN2A-CDKN2B, IDE-KIF11-HHEX, IGF2BP2 and SLC30A8 are associated with type 2 diabetes in a Chinese population. *PLoS ONE* 4:e7643. <https://doi.org/10.1371/journal.pone.0007643>
- Hu Y, Flockhart I, Vinayagam A et al (2011) An integrative approach to ortholog prediction for disease-focused and other functional studies. *BMC Bioinform* 12:357. <https://doi.org/10.1186/1471-2105-12-357>

- Jennings BH (2011) *Drosophila*: a versatile model in biology & medicine. *Mater Today* 14:190–195. [https://doi.org/10.1016/S1369-7021\(11\)70113-4](https://doi.org/10.1016/S1369-7021(11)70113-4)
- King AJ (2012) The use of animal models in diabetes research. *Br J Pharmacol* 166:877–894. <https://doi.org/10.1111/1j.1476-5381.2012.01911.x>
- Klimentidis YC, Abrams M, Wang J et al (2011) Natural selection at genomic regions associated with obesity and type-2 diabetes: East Asians and sub-Saharan Africans exhibit high levels of differentiation at type-2 diabetes regions. *Hum Genet* 129:407–418. <https://doi.org/10.1007/s00439-010-0935-z>
- Koenker R, Portnoy S, Ng PT, et al (2018) Package ‘quantreg’
- König B (2001) Natural Selection. In: Smelser NJ, Baltes PB (eds) *International encyclopedia of the social & behavioral sciences*. Pergamon, Oxford, pp 10392–10398
- Kosiol C, Vinař T, da Fonseca RR et al (2008) Patterns of positive selection in six mammalian genomes. *PLoS Genet* 4:e1000144. <https://doi.org/10.1371/journal.pgen.1000144>
- Krieger MJB, Ross KG (2002) Identification of a major gene regulating complex social behavior. *Science* 295:328–332. <https://doi.org/10.1126/science.1065247>
- Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 26:283–291
- Lee G, Park JH (2004) Hemolymph sugar homeostasis and starvation-induced hyperactivity affected by genetic manipulations of the adipokinetic hormone-encoding gene in *Drosophila melanogaster*. *Genetics* 167:311–323
- Lefebvre T, Stanhope MJ (2009) Pervasive, genome-wide positive selection leading to functional divergence in the bacterial genus *Campylobacter*. *Genome Res* 19:1224–1232. <https://doi.org/10.1101/gr.089250.108>
- Lemkul JA (2015) GROMACS tutorial: KALP15 in DPPC. Retrieved February
- Little BB, Reyes MEP, Malina RM (2017) Natural selection and type 2 diabetes-associated mortality in an isolated indigenous community in the valley of Oaxaca, southern Mexico. *Am J Phys Anthropol* 162:561–572. <https://doi.org/10.1002/ajpa.23139>
- Liu Y, He W, Long J et al (2013) Natural selection and functional diversification of the epidermal growth factor receptor EGFR family in vertebrates. *Genomics* 101:318–325. <https://doi.org/10.1016/j.ygeno.2013.03.001>
- Lynn DJ, Freeman AR, Murray C, Bradley DG (2005) A genomics approach to the detection of positive selection in cattle. *Genetics* 170:1189–1196. <https://doi.org/10.1534/genetics.104.039040>
- Maurer KJ, Quimby FW (2015) Chapter 34: animal models in biomedical research. In: Fox JG, Anderson LC, Otto GM, et al. (eds) *Laboratory animal medicine*, 3rd edn. Academic Press, Boston, pp 1497–1534
- Morgan CC, Shakya K, Webb A et al (2012) Colon cancer associated genes exhibit signatures of positive selection at functionally significant positions. *BMC Evol Biol* 12:114. <https://doi.org/10.1186/1471-2148-12-114>
- Murillo-Maldonado JM, Riesgo-Escovar JR (2017) Development and diabetes on the fly. *Mech Dev* 144:150–155. <https://doi.org/10.1016/j.mod.2016.09.004>
- Musselman LP, Fink JL, Narzinski K et al (2011) A high-sugar diet produces obesity and insulin resistance in wild-type *Drosophila*. *Dis Model Mech* 4:842–849. <https://doi.org/10.1242/dmm.007948>
- Nesse RM, Stearns SC (2008) The great opportunity: Evolutionary applications to medicine and public health. *Evol Appl* 1:28–48. <https://doi.org/10.1111/j.1752-4571.2007.00006.x>
- Ngai J, Dowling MM, Buck L et al (1993) The family of genes encoding odorant receptors in the channel catfish. *Cell* 72:657–666
- O’Halloran TV, Kebede M, Philips SJ, Attie AD (2013) Zinc, insulin, and the liver: a ménage à trois. *J Clin Invest* 123:4136–4139. <https://doi.org/10.1172/JCI12325>
- Pound LD, Oeser JK, O’Brien TP et al (2013) G6PC2: a negative regulator of basal glucose-stimulated insulin secretion. *Diabetes* 62:1547–1556. <https://doi.org/10.2337/db12-1067>
- RC Team (2014) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, p 2014
- Rees DA, Alcolado JC (2005) Animal models of diabetes mellitus. *Diabet Med* 22:359–370. <https://doi.org/10.1111/j.1464-5491.2005.01499.x>
- Rutter GA, Chimienti F (2015) SLC30A8 mutations in type 2 diabetes. *Diabetologia* 58:31–36. <https://doi.org/10.1007/s00125-014-3405-7>
- Ségurel L, Austerlitz F, Toupance B et al (2013) Positive selection of protective variants for type 2 diabetes from the Neolithic onward: a case study in Central Asia. *Eur J Hum Genet* 21:1146–1151. <https://doi.org/10.1038/ejhg.2012.295>
- Skyler JS, Bakris GL, Bonifacio E et al (2017) Differentiation of diabetes by pathophysiology, natural history, and prognosis. *Diabetes* 66:241–255. <https://doi.org/10.2337/db16-0806>
- Spires D, Manis AD, Staruschenko A (2019) Ion channels and transporters in diabetic kidney disease. *Curr Top Membr* 83:353–396. <https://doi.org/10.1016/bs.ctm.2019.01.001>
- Swanson WJ, Nielsen R, Yang Q (2003) Pervasive adaptive evolution in mammalian fertilization proteins. *Mol Biol Evol* 20:18–20. <https://doi.org/10.1093/oxfordjournals.molbev.a004233>
- Szklarczyk D, Morris JH, Cook H et al (2017) The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Res* 45:D362–D368. <https://doi.org/10.1093/nar/gkw937>
- Teng L, Fan X, Xu D et al (2017) Identification of genes under positive selection reveals differences in evolutionary adaptation between brown-algal species. *Front Plant Sci*. <https://doi.org/10.3389/fpls.2017.01429>
- Thirlaway K, Davies L (2001) Lifestyle responses to genetic susceptibility to type 2 diabetes. Wiley, Hoboken
- Tiwari P (2015) Recent trends in therapeutic approaches for diabetes management: a comprehensive update. *J Diabetes Res*. <https://www.hindawi.com/journals/jdr/2015/340838/>. Accessed 23 Jan 2019
- Turner PJ (2005) XMGRACE, Version 5.1. 19. Center for Coastal and Land-Margin Research. Oregon Graduate Institute of Science and Technology, Beaverton, OR
- Vishnoi A, Kryazhimskiy S, Bazykin GA et al (2010) Young proteins experience more variable selection pressures than old proteins. *Genome Res* 20:1574–1581. <https://doi.org/10.1101/gr.109595.110>
- Wagner A (2007) Rapid detection of positive selection in genes and genomes through variation clusters. *Genetics* 176:2451–2463. <https://doi.org/10.1534/genetics.107.074732>
- Walker BR, Colledge NR (2013) *Davidson’s Principles and Practice of Medicine E-Book*. Elsevier Health Sciences, Amsterdam
- Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res* 35:W407–W410. <https://doi.org/10.1093/nar/gkm290>
- Willett CS (2000) Evidence for directional selection acting on pheromone-binding proteins in the genus *Choristoneura*. *Mol Biol Evol* 17:553–562. <https://doi.org/10.1093/oxfordjournals.molbev.a026335>
- Wolf MG, Hoefling M, Aponte-Santamaría C et al (2010) g\_memb: efficient insertion of a membrane protein into an equilibrated lipid bilayer with minimal perturbation. *J Comput Chem* 31:2169–2174. <https://doi.org/10.1002/jcc.21507>

- Wong A, Turchin MC, Wolfner MF, Aquadro CF (2008) Evidence for positive selection on *Drosophila melanogaster* seminal fluid protease homologs. *Mol Biol Evol* 25:497–506. <https://doi.org/10.1093/molbev/msm270>
- Yang W, Bielawski JP, Yang Z (2003) Widespread adaptive evolution in the human immunodeficiency virus type 1 genome. *J Mol Evol* 57:212–221. <https://doi.org/10.1007/s00239-003-2467-9>
- Yang Z (2006) Computational molecular evolution. OUP, Oxford
- Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24:1586–1591. <https://doi.org/10.1093/molbev/msm088>
- Yang Z, Bielawski JP (2000) Statistical methods for detecting molecular adaptation. *Trends Ecol Evol* 15:496–503
- Yang Z, Swanson WJ (2002) Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. *Mol Biol Evol* 19:49–57. <https://doi.org/10.1093/oxfordjournals.molbev.a003981>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.