UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
INSTITUTO DE BIOCIÊNCIAS

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS
(Biologia Celular e Molecular)

UNDERSTANDING THE EVOLUTION OF SATELLITE DNAS AT INTRASPECIFIC LEVEL BY ANALYSIS OF THE LIBRARY IN THE HOLOCENTRIC PEST APHID *Acyrthosiphon pisum* (HEMIPTERA)

LUCAS ALBUQUERQUE DOS SANTOS

Novembro, 2021

LUCAS ALBUQUERQUE DOS SANTOS

UNDERSTANDING THE EVOLUTION OF SATELLITE DNAS AT INTRASPECIFIC LEVEL BY ANALYSIS OF THE LIBRARY IN THE HOLOCENTRIC PEST APHID *Acyrthosiphon pisum* (HEMIPTERA)

Orientador: Prof. Dr. Diogo C. Cabral-de-Mello

Co-orientador(a): Dr. Octavio M. Palacios Gimenez

Texto apresentado ao Instituto de Biociências do Câmpus de Rio Claro, Universidade Estadual Paulista, como parte dos requisitos para qualificação no programa de pós-graduação em Ciências Biológicas, Área de Concentração: Biologia Celular e Molecular.

Rio Claro - SP

2021

A345u

Albuquerque-dos-Santos, Lucas
Understanding the evolution of satellite DNAs at intraspecific level by analysis of the library in the holocentric pest aphid Acyrthosiphon pisum (Hemiptera) / Lucas Albuquerque-dos-Santos. -- Rio Claro, 2022
48 p. : il., tabs., fotos

Dissertação (mestrado) - Universidade Estadual Paulista (Unesp), Instituto de Biociências, Rio Claro
Orientador: Diogo C. Cabral-de-Mello
Coorientador: Octavio M. Palacios-Gimenez

1. Genética. 2. DNA satélite. 3. Afídeos. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca do Instituto de Biociências, Rio Claro. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Rio Claro

CERTIFICADO DE APROVAÇÃO

TÍTULO DA DISSERTAÇÃO: UNDERSTANDING THE EVOLUTION OF SATELLITE DNAS AT INTRASPECIFIC LEVEL BY ANALYSIS OF THE LIBRARY IN THE HOLOCENTRIC PEST APHID Acyrthosiphon pisum (HEMIPTERA)

AUTOR: LUCAS ALBUQUERQUE DOS SANTOS
ORIENTADOR: DIOGO CAVALCANTI CABRAL DE MELLO
COORIENTADOR: OCTAVIO MANUEL PALACIOS GIMÉNEZ

Aprovado como parte das exigências para obtenção do Título de Mestre em CIÊNCIAS BIOLÓGICAS (BIOLOGIA CELULAR E MOLECULAR), pela Comissão Examinadora:

Diogo C. Cabral de Mello

Prof. Dr. DIOGO CAVALCANTI CABRAL DE MELLO (Participaçao Virtual)
Departamento de Biologia Geral e Aplicada / Unesp - Instituto de Biociencias - Rio Claro

Profa. Dra. PATRICIA PASQUALI PARISE MALTEMPI (Participaçao Virtual)
Departamento de Biologia Geral e Aplicada / Unesp - Instituto de Biociencias - Rio Claro

Prof. Dr. ELIO RODRIGO CASTRO (Participaçao Virtual)
Laboratorio de Genética Evolutiva / Universidad Nacional de Misiones

Rio Claro, 29 de novembro de 2021

Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001. Apoio pelo qual devo minha eterna gratidão. Agradeço também pelo fomento do projeto "Investigando a evolução de DNAs satélites em cromossomos holocêntricos de insetos pragas" através do processo nº 19/19069-7, Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), e também agradeço ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Agradeço à Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP) e sua comunidade, especialmente no câmpus de Rio Claro, e ao Departamento de Biologia Celular e Molecular, por abrigar esse trabalho e minha educação, e prover as condições necessárias para o andamento do projeto. Agradeço a todos que lutam pela resiliência dessas instituições. Agradeço ao Professor Doutor Diogo Cavalcanti Cabral de Mello pela orientação e ao Doutor Octavio Manuel Palacios Gimenez, pela coorientação do trabalho. Agradeço aos meus amigos de laboratório e suas famílias, principalmente Diogo Milani, Emiliano Martí, Beatriz Ferretti, Vanessa Bardella, Ana Elisa Gasparotto e Gerson Mello. Agradeço a colaboração da Professora Doutora Jenn Brisson da Universidade de Rochester e seu aluno de doutorado Omid Ziabari. Agradeço aos cientistas que serviram de fonte e inspiração.

Dedico este trabalho à minha família e às vítimas da pandemia do coronavírus, que seu legado de alguma maneira seja a luz que nos resgatará da sombra que nos tornamos.

Summary

Resumo

A evolução da biblioteca de DNA satélite é um campo de pesquisa muito relevante, pois a composição e o arranjo dessas sequências têm impacto na arquitetura do genoma e seu papel funcional ainda é pouco conhecido. Existe uma dificuldade em estudar esse tipo de material genético. Devido às limitações tecnológicas, o sequenciamento genômico moderno tende a subestimar a quantidade de DNA repetitivo e isso desafia uma análise mais precisa. No entanto, um excelente trabalho tem sido desenvolvido com a disseminação de sequenciamentos de alta cobertura, permitindo uma investigação mais aprofundada das espécies e comparação entre os grupos. Este trabalho propõe uma contribuição envolvendo a investigação da hipótese da biblioteca de DNA satélite e sua evolução em concerto em nível intraespecífico usando 16 populações da espécie de pulgão praga agrícola *Acyrthosiphon pisum*. Esta espécie, que carrega cromossomos holocêntricos, é caracterizada por sua especialização em diferentes plantas hospedeiras, tornando-se um modelo chave para análise populacional e diferenciação genética. Aqui, usamos a plataforma RepeatExplorer para prospectar famílias de DNA de satélite em genomas de *A. pisum* disponíveis no banco de dados NCBI. Construímos uma biblioteca com sequências confiáveis e comparamos a abundância e divergência dessa biblioteca em cada genoma e também em 3 outras espécies de afídeos com o software RepeatMasker. Descobrimos que a maioria das famílias de satélites são compartilhadas entre os biótipos e espécies irmãs, mas com variedade marcada em abundância e homogeneização. Também analisamos a abundância desta biblioteca em montagens cromossômicas de leitura longa de uma linhagem e encontramos um maior acúmulo de famílias de satélites no cromossomo X. Nossos resultados contribuem com a literatura recente no entendimento da evolução da biblioteca de DNA satélite em geral e neste importante modelo.

Palavras-chave: DNA repetitivo, diferenciação genômica, cromossomo holocêntrico, bioinformática

Abstract

The evolution of the satellite DNA library is a very relevant research field, as the composition and arrangement of these sequences have an impact on the genome's architecture and their functional role is still poorly understood. There is a difficulty in studying this type of genetic material. Due to technological limitations, modern genomic sequencing tends to underestimate the amount of repetitive DNA and this challenges more accurate analysis. However, excellent work has been developed with the dissemination of high coverage sequencing, allowing a deeper investigation into the species and comparison between groups. This work proposes a contribution involving the investigation of the satellite DNA library hypothesis and its evolution in concert at an intraspecific level using 16 populations of the agricultural pest aphid species *Acyrthosiphon pisum*. This species, which carries interesting holocentric chromosomes, is characterized by its specialization to different host plants, making it a key model for population analysis and genetic differentiation. Here we use the RepeatExplorer platform to prospect satellite DNA families in *A. pisum* genomes available in the NCBI database. We built a library with reliable sequences and compared the abundance and divergence of that library in each genome and also in 3 other aphid species with the RepeatMasker software. We found that most families of satellites are shared between the biotypes and the sister species, but with marked variety in abundance and homogenization. We also analyzed the abundance of this library in long read chromosomal assemblies of a biotype and found a greater accumulation of satellite families on the X chromosome. Our results contribute to recent literature in understanding the evolution of the satellite DNA library in general and in this important model.

Key-words: repetitive DNA, genomic differentiation, holocentric chromosome, bioinformatics

1 Introduction

1.1 Satellite DNAs

The linear arrangement of DNA directly impacts the structural and spatial organization of the genetic material and, consequently, its evolutionary dynamics (LOUZADA et al., 2020). The genomes of eukaryotic organisms are not composed only by unique structural genes, but they also contain DNA sequences that are repeated from two to thousands of copies, organized in tandem or dispersed through the genome, whose composition differs greatly from the rest of the genome (GARRIDO-RAMOS, 2017). One of these types of repetitive DNAs is called satellite DNA or satDNA (THAKUR et al., 2021).

Normally, satDNAs together with the other class of dispersed repetitive elements (the transposable elements-TEs) can make up more than 50% of the genome content (CHARLESWORTH et al., 1994). The differences between them and the rest of the genetic material and its tandem organization allow the satDNA repeat fractions to be centrifuged and separated from the rest of the DNA by gradient of density (PLOHL et al., 2012). These DNA fragments are often found in heterochromatin, especially in centromeric and terminal regions of chromosomes, although intercalary and euchromatic satDNAs have also been reported (PARDUE and GALL, 1970; JONES, 1970; PALACIOS-GIMENEZ et al., 2020; SPROUL et al., 2020).

Despite decades of intensive research, understanding of the functional role of this type of DNA is still limited. Due to the size of the fragments generated by the widely disseminated sequencing (with short fragments), there is a challenge to accurately quantify the abundance of repetitive sequences in the genome (LOWER et al., 2018). This challenge is being addressed through the constant development of softwares for this type of analysis and by distinct methods of sequencing that generates longer fragments, which allows observing a larger linear arrangement of the DNA sequences in the genomes (VONDRAK et al., 2019).

In the 1970s it was proposed that repetitive DNAs were a useless portion of the genome, a type of genetic waste dump (OHNO, 1972), as they are often associated with the constitutive heterochromatin which was poor in genes, genetics activity and remain condensed during cell division (ORGEL and CRICK, 1980). At the same time, studies have investigated the role of repetitive DNAs in sex chromosome differentiation (SINGH et al., 1976), in the macro-structural organization of the genome (HATCH et al., 1976) and in speciation (MAZRIMAS et al., 1972). Over the decades, with the development of technologies and the increase in resolution of the analysis, it has been proposed that a rapid change in the composition of the satDNAs in an organism can lead to a key speciation process by an adaptive pressure on the proteins involved in heterochromatinization and in centromeric regions, important for cell division (HENIKOFF et al., 2001), including in humans (SCHUELER et al., 2001). Recently, it was demonstrated in *Drosophila* hybrids that differences in composition of pericentromeric satDNAs disrupt directly the organization of the chromocenters, cause formation of micronuclei and atrophy of tissues, promoting hybrid incompatibility (HI), that ultimately could be causes of reproductive isolation (JAGANNATHAN and YAMASHITA, 2021).

The satDNAs are a quite dynamic portion of genomes and even closely related species can diverge significantly in the composition of these repetitive DNAs. Although, some cases of coherence in distantly related species were observed, which makes their analysis of interest in understanding the evolution and speciation of organisms (HOLMQUIST et al., 1979). The satDNA differs in nucleotide composition, sequence complexity, repeat unit length, and abundance, besides specific organization on chromosomes. Only two characteristics are shared between the satDNAs, they are (i) their capacity to constitute ordered arranged repetitions in tandem and (ii) their frequent association with the constitutive heterochromatin (PLOHL, 2012). It has been

postulated that satDNAs evolve in concert (DOVER et al, 1982; LIAO, 1999), i.e. concerted evolution, resulting in the homogenization of repeat changes in the genome caused by mutations and their subsequent fixation on reproductive members in the population (URGAKOVIC et al., 2002) by a molecular drive mechanism (DOVER, 1986). Phenomena such as unequal crossing over (STURTEVANT, 1925) and gene conversion (HOLLIDAY, 1964) are responsible for spreading new mutations between members of repetitive families within a genome (URGAKOVIC et al., 2002). Unequal crossing over is also responsible for the change in the number of repeated copies and size of the satellite array (SMITH, 1976). Theoretical studies on satDNA dynamics explain their reduction in the genome by unequal crossing over, demonstrating an inverse correlation between the rate of this event and the preservation of the satellite over time (STEPHAN, 1986). The number of copies can also increase in a short period of time by mechanisms such as replication slippage (TACHIDA et al., 1992), rolling circle replication (STENGER et al., 1991), mechanisms similar to gene conversion and other processes still unknown (CHARLESWORTH et al., 1994). Copy number and nucleotide sequence can be independently analyzed. The number of copies of satDNA varies widely between related species (KING et al., 1995; VERSHININ et al., 1996), as well as the location in chromosomes (ROSS et al. 1997). Regarding the sequence, for example, a considerable fluctuation of distinct satellites was found in the genome of some types of cattle (NIJMAN and LENSTRA, 2001) despite the short time of divergence between species. In insects, Palacios-Gimenez et al. (2020a) observed a distinguished proliferation of satDNA families in races of the grasshopper *Vandiemenella viatica*.

Fry and Salser (1977) proposed that related species can share a common satDNA library, varying in specific number of copies, this is named the library hypothesis. This was experimentally shown in multiple groups of species, like in beetles (PONS et al.,

2004), grasshoppers (PALACIOS-GIMENEZ et al., 2020b), mammals (SMALEC et al., 2019), plants (QUESADA DEL BOSQUE et al., 2011), among others. The number of copies of a satellite DNA can be directly linked to the evolution of chromosomes in a species, and consequently to an eventual speciation (SLAMOVITS et al., 2001; MESTROVIC et al., 1998). The nucleotide sequences present an internal variability, which depends on the ratio between the mutation rate and the fixation of the sequences (DOVER, 1986). Thus, each satDNA unit can be considered an independent evolutionary unit, and unequally distributed among chromosomes, forming specific subfamilies without necessarily changing the quantity (ARNASON et al., 1992). The study of satDNAs in natural populations contributes to the understanding of evolutionary forces acting on the dynamics of speciation, whether by genome architecture or possible functions yet to be elucidated in different groups of eukaryotes (GARRIDO-RAMOS, 2015; PALOMEQUE et al., 2008; YANG et al., 2020; VOURC'H et al., 2011; PEZER et al., 2012).

1.2 The aphid *Acyrthosiphon pisum*: an emergent model for genomic studies

Aphids are plant sap-sucking insects of great agricultural and entomological interest due to their direct parasitic action on crops and their role as a vector of viruses that can cause losses in agricultural productivity (VAN EMDEN and HARRINGTON, 2007; DESHOUX et al., 2020). The aphid *Acyrthosiphon pisum* (HARRIS, 1776) belongs to the Aphidoidea family, of the order Hemiptera. This species forms a complex of sympatric populations that feed on more than 20 legume genera, including for example alfalfa, beans and peas (PECCOUD and SIMON 2010). Each race or biotype of the species is adapted to one or a few legume species and are genetically different from each other (VIA, 1991; FRANTZ et al., 2009; FERRARI et al., 2008, 2012), due to distinct selective pressures, caused by, for example, by defense mechanisms developed

by host plants (SCHWARZKOPF, 2013). It is a species with original distribution in the Palearctic region, but which was possibly introduced multiple times in the new world in the last 200 years (BLACKMAN, 1981). Introduced aphids showed a high frequency of some similar genotypes due to their parthenogenetic reproduction (MORALES-HOJAS, 2017; LYNCH, 1984), a condition that in an environment, even if distant, but with attributes similar to their original habitat, can result in the occurrence of what was called "superclone" by Harrison et al. (2011). In these superclones the pre- adaptations and neutral mutations become positive and are widely diffused in the new environment in which they were introduced, causing a rapid local adaptation by parthenogenetic clonal lineages (LOXDALE et al., 2018). Figueroa et al. (2018) found in Chile three superclones that were present with the same ecological specialization (biotypes) found in France (PECCOUD et al., 2008). These Chilean genotypes were also more recently studied by Serteyn et al. (2021).

Chromosomes can undergo rearrangements by various mechanisms, and this dynamic has evolutionary implications (EICHLER and SANKOFF, 2003). At the cytogenetic level, aphids are a very interesting model for having holocentric chromosomes, exhibiting centromeric activity spread along the entire axis of the chromosomes, that are prone to tolerate more the chromosomal rearrangements (HUGHES-SCHRADER and SHRADER, 1961; BLACKMAN et al., 1987; MELTERS et al. 2012). Other groups of non-related animals and plants also have holocentric chromosomes, and so far, it is believed that they were adaptive convergences (DERNBURG, 2001; MANDRIOLI and MANICARDI, 2012; MELTERS et al., 2012; HECKMANN and HOUBEN, 2013). In *A. pisum*, the karyotype is composed of 2n = 8, XX (female) and 2n = 7, X0 (male), with the X chromosome as the largest element (BIZZARO et al. 2000). For the organization of repetitive DNAs in the chromosomes of the species, there is information obtained by banding techniques (C banding and

staining with base-specific fluorochromes) and by mapping through Fluorescent *in situ* Hybridization (FISH) of rDNA sequences (5S and 28S), H3 histone gene, telomeric sequence (TTAGG)*n* and one retrotransposon. The constitutive heterochromatin, which is rich in G+C, is concentrated in one terminal region of the X chromosome that also harbors the 28S rDNA cluster. The 5S rDNA is also located on the X chromosome in interstitial regions, while the histone H3 gene is on the longest autosome chromosome (BIZZARO et al. 2000; MANDRIOLI and MANICARDI, 2013). The telomeric sequence and the retrotransposon TRAS are located in the terminal region of chromosomes (BIZZARO et al. 2000; MONTI et al. 2013).

The genomic sequencing of *A. pisum* (THE INTERNATIONAL APHID GENOMICS CONSORTIUM, 2010) allowed significant advances in the understanding of biological characteristics of the species. Furthermore, recently Mandrioli and collaborators (2017) elaborated a comparative gene map with genetic elements that have been well studied in fruit flies (*Drosophila* spp., Diptera) to understand the evolution of their chromosomes at the gene level in *A. pisum*, and observed that traits (80 genes) of Muller elements are present in *A. pisum* even though they diverged for at least 330 million years ago (HEDGES et al., 2015).

Fernández et al. (2020) found a relationship between *A. pisum* gene duplication and selection for neo-functionalization of paralogous genes, modeling the recent evolutionary dynamics of the genome and adaptive plasticity to host plants. This gene duplication and neofunctionalization has also been studied in aphids recently by Julca and collaborators (2020). In 2019, Li and colleagues published an analysis of gene families in new *A. pisum* genome assemblies at the chromosomal level. This assembly was reconstructed by Mathers et al. (2020) and compared to other aphid assemblies at the chromosomal level, revealing that *A. pisum* autosomal chromosomes have undergone intense reorganization over the past 30 million years (also noted by Li et al.,

2020). While the gene content of the X chromosome was slightly altered, it also presents a significant enrichment of repetitive elements and impoverishment of genes in relation to the autosomes.

Genomic studies have been conducted using *A. pisum* as a model, but it is necessary to improve the knowledge regarding the structure, organization and evolution of repetitive DNAs, including the satDNAs. In the present work we advance in the understanding of the satDNA library in *A. pisum*, using the species as model to understand the patterns of evolution of satDNA library at intraspecies level.

2 Research Aims and Objectives

2.1 Main Objective

Study the library hypothesis at intraspecific level, using as model distinct biotypes of the aphid *Acyrthosiphon pisum*, a species with holocentric chromosomes.

2.2 Specific Objectives

-To prospect and characterize at molecular point of view the satDNAs from distinct biotypes of *A. pisum*;

- Analyze the patterns of evolution of satDNAs between biotypes;

- Study the patterns of organization and evolution of satDNAs between chromosomes;

- To map on chromosomes the most abundant satDNAs from *Medicago sativa* biotype.

3 Results

The results of this dissertation are presented as a manuscript in elaboration stage entitled "Understanding the evolution of satellite DNAs at intraspecific level by analysis of the library in the holocentric pest aphid *Acyrthosiphon pisum* (Hemiptera)". This

manuscript will be submitted to the journal Journal of Molecular Evolution, ISSN 1432-1432.

3.1 Manuscript

Understanding the evolution of satellite DNAs at intraspecific level by analysis of the library in the holocentric pest aphid *Acyrthosiphon pisum* (Hemiptera)

Lucas Albuquerque[1], Diogo Milani[1], Emiliano Martí[1], Ana BSM Ferretti[1], Octavio M Placios-Gimenez[2,3], Omid Saleh Ziabari[4], Jennifer Brisson[4], Diogo C Cabral-de-Mello[1,*]

[1]Departamento de Biologia Geral e Aplicada, UNESP - Univ Estadual Paulista, Instituto de Biociências/IB, Rio Claro, São Paulo, Brazil
[2]Department of Organismal Biology – Systematic Biology, Evolutionary Biology Centre, Uppsala University, SE-752 36 Uppsala, Sweden
[3]Institute of Ecology and Evolution, Friedrich Schiller University Jena, 07743 Jena, Germany
[4]Department of Biology, University of Rochester, Rochester, NY 14627, USA

**Correspondence:* Diogo C. Cabral-de-Mello
UNESP - Univ Estadual Paulista, Instituto de Biociências/IB, Departamento de Biologia Geral e Aplicada, 13506-900 Rio Claro, SP, Brazil Phone/Fax: 55 19 35264152.
E-mail: cabral.mello@unesp.br (D.C.C.-d-M.)

Short running title: Intraspecific satDNA evolution in *Acyrthosiphon pisum*

Abstract

The evolution of the satellite DNA library is a very relevant research field, as the composition and arrangement of these sequences have an impact on the genome's architecture and their functional role is still poorly understood. However, excellent contributions have been developed with the dissemination of high coverage sequencing, allowing a deeper investigation into the species and comparison between groups. This work proposes a contribution involving the investigation of the satellite DNA library hypothesis and its evolution in concert at an intraspecific level using 16 populations of the agricultural pest aphid species *Acyrthosiphon pisum*. This species, which carries interesting holocentric chromosomes, is characterized by its specialization to different host plants, making it a key model for population analysis and genetic differentiation. We used the RepeatExplorer platform to prospect satellite DNA families in *A. pisum* genomes available in the NCBI database. We built a library with reliable sequences and compared the abundance and divergence of that library in each genome and also in 3 other aphid species with the RepeatMasker software. We found that most families of satellites are shared between the biotypes and the sister species, but with marked variety in abundance and homogenization. We also analyzed the abundance of this library in long read chromosomal assemblies of a biotype and found a greater accumulation of satellite families on the X chromosome. Our results contribute to recent literature in understanding the evolution of the satellite DNA library in general and in this important model.

Key-words: repetitive DNA, genomic differentiation, holocentric chromosome, bioinformatics

Introduction

The genomes of eukaryotic organisms are plenty of DNA sequences that could be repeated from hundreds to thousands of copies. These sequences are dispersed throughout genomes in multiple loci or they are in tandem clustered occupying specific regions, differing largely in composition from the rest of the genome (Garrido-Ramos 2017; Louzada et al. 2020; Thakur et al. 2021). One of these types of repetitive DNAs are the satellite DNA (satDNA) that normally together with dispersed repetitive elements (transposons and retrotransposons) can make up more than 50% of the genome of organisms (Charlesworth et al. 1994; López-Flores and Garrido-Ramos 2012; Garrido-Ramos 2017; Thakur et al. 2021). The satDNAs are quite dynamic genome elements, and they differ in base-pairs nucleotide sequence, sequence complexity, repeat unit length, and abundance. Only two general characteristics are more commonly conserved among satDNAs, (i) the ability to constitute ordered in tandem repetitions arrangements (ii) and the frequent association with the constitutive heterochromatin, frequently on centromeres and telomeres (Plohl et al. 2012; Thakur et al. 2021).

The high dynamism of satDNAs could be evidenced also in closely related species that can diverge significantly in the composition of these repetitive DNAs (Garrido-Ramos 2017), although in some phylogenetic groups, some coherence was maintained (*e.g.* Plohl et al. 2010; Escudeiro et al. 2019; Smalec et al. 2019; Palacios-Gimenez et al. 2020). This high variation in satDNA among related species, including sequence variation and copy number, can be attributed to relaxed evolution, putatively associated to the lack of function (Lower et al. 2018).

Distinct satDNA families could be found in a genome, i.e. the satDNA library. According to the library hypothesis, the set of satDNAs shared among species originates in the common ancestor and during the evolutionary time this library experiences differential amplification or contraction after speciation processes (Salser et al. 1976;

Fry and Salser 1977; Mestrovic et al. 1998, GONZÁLES et al. 2020). The evolutionary model to explain the evolution of satDNAs is the concerted evolution, that led to the homogenization of the satDNA repeats (Dover 1986; Elder and Turner 1995; Ugarkovic and Plohl 2002; Plohl et al. 2008; Garrido-Ramos 2017). The molecular drive is responsible for the concerted evolution and some molecular mechanisms are the main factors implicated in the process, like unequal crossover, slippage rolling-circle replication, gene conversion, transpositions and gene conversion (Dover 1982; 2002; Walsh 1987; Shi et al. 2010). The mutational processes followed by concerted evolution is the explanation of the high divergence between satDNA libraries among distinct species, which can cause the fixation of new variants in reproductive groups. The rapid differentiation of satDNAs, due to their intense dynamism accounted by the mechanisms above, can have impact in reproductive isolation and ultimately contribute to speciation (Bachmann et al. 1989; Mestrovic et al. 1998; Malik et al. 2001; Ferree and Prasad 2012; Jagannathan and Yamashita 2021).

Although profusely studied, proving to be a quite intriguing part of the genomes, the evolution of satDNAs were not well studied at intraspecies level. Moreover, most evolutionary models emerged from empirical analysis of species with monocentric chromosomes (localized centromeres), and only few data were explored in holocentric related species (with diffuse centromere), as in plants from genus *Rhynchospora* (Ribeiro et al. 2017), Triatominae bugs (Pita et al. 2017), Crambidae moths (Cabral-de-Mello et al. 2021) and *Caenorhabditis* nematodes (Subirana et al. 2015). In this way, aiming to understand the evolution of satDNA library at intraspecific level and on holocentric chromosomes, here we studied the satDNAs comprising the library of the pest aphid model species *Acyrthosiphon pisum* (Hemiptera), the pea aphid. This aphid is an emerging model for diverse analysis, including genomic and evolutionary studies (Brisson and Stern 2006). Thus, we took the advantage of genomic resources from 16

biotypes, that are adapted to one or a few legume species and are genetically different from each other (Ferrari et al. 2008; 2012) to compare the patterns of organization and differentiation of satDNAs, allowing to test the hypothesis postulated for satDNAs evolution at single species level. Moreover, we studied details about evolution of satDNA families between chromosomes by analysis of the assembled genome.

Material and Methods

Genome datasets

We selected 19 genomes sequenced by Illumina Whole Genome Sequencing (WGS) pair-end short reads (100bp-150bp) from Sequence Reads Achieve of the NCBI www.ncbi.nlm.nih.gov/sra (Supplementary Table 1). They corresponded to 16 populations, including 11 biotypes associated to distinct plants of *Acyrthosiphon pisum* (collection access numbers PRJNA385905 and PRJNA454786) deposited by the French *Institut National de la Recherche Agronomique* (INRA) with average of 25 pooled individuals per population. These samples include populations hosted in the same species of plant, plants from the same genus, and plants of different genus.

Moreover, for comparative analysis we selected other three WGS from sister species, including *Acyrthosiphon caraganae* (SRX6674381), *Aphis craccivora* (SRX2888378), and *Myzus persicae* (SRX6674382) that were selected bearing in mind the relationship between the species based on the recent phylogeny (Mathers et al. 2021). These genomes were used for satDNAs prospection, characterization and analysis to understand patterns of intraspecific evolution of satDNAs.

Additionally, we selected the chromosome assemblies for the four chromosomes (Chr-A1, Chr-A2, Chr-A3 and Chr-X) of *A. pisum* (Li et al. 2019) available on the NCBI (bioproject PRJNA496478) to conduct an intragenomic satDNA analysis among chromosomes for abundance and divergence for each repeat. Finally, we analyzed the

genome assembly (Mathers et al. 2021) to verify more precisely the organization of the sat DNAs in *A. pisum* chromosomes.

satDNA library analysis

As the genomic data were large, first we randomly sampled 10 million pair-end reads per genome in the Galaxy online platform (Afgan et al. 2016) using the public server usegalaxy.org. The raw reads were pre-processed for trimming, filtering and read interlacing on the Galaxy online platform with default parameters, quality cutoff = 10 and percent above cutoff = 95. The Tandem Repeat Analyzer, *i.e.* TAREAN (Novák et al. 2017), was executed in each of the 16 *A. pisum* samples to the satDNA families identification. To be more confident in identification of satDNAs we considered for further analysis only satDNAs that were identified at least one time (in one of the populations) as "high confidence" by TAREAN.

The monomers of each satDNA were grouped in the Geneious 4.8.5 software (http://www.geneious.com) for homology analysis among them based on the family classification proposed by Ruiz-Ruano et al (2016) to build the *A. pisum* satDNA library containing consensus sequences. The generated library was used as custom database to evaluate and compare the abundance and divergence of satDNA families in the 16 biotypes and the other four sister aphid species using the RepeatMasker Open-4.0 software (http://www.repeatmasker.org) considering the number of nucleotides of each sample for the abundance calculation. Additionally, we also calculated using the same custom satDNA database the abundance and divergence of satDNAs between the assembled chromosomes also using RepeatMasker. The genomic abundance for each satDNA family is calculated by the proportion of nucleotides aligned with the consensus sequence divided by the sequenced library size and the script calcDivergenceFromAlign.pl estimate the average Kimura2-parameter distances

(Kimura, 1980) for each satDNA family. The satDNA were named as ApiSat and ordered according to the abundance decreasing followed by base pairs monomer size, according to Ruiz-Ruano et al. (2016).

The satDNA collection was masked against the assembled genome with RepeatMasker to define the position of each satDNA along the chromosomes. Then, we plotted graphical images with satDNAs positions using the KaryotypeR R package (Gel and Serra 2017).

Fluorescence in situ hybridization (FISH)

For physical chromosomal characterization of satDNAs we selected five families based on their abundance, divergence and genome organization revealed by computational analysis (see results). The probes for satDNAs were labeled directly during their synthesis with Biotin at 5'end. The heterologous probe for 18S rDNA was amplified from the genome of *Diatraea saccharalis* using the primers Sca18SF (5'CCC CGT AAT CGGAAT GAG TA), Sca18SR (5'GAG GTT TCC CGT GTTGAG TC) (Cabral-de-Mello et al. 2010) and labeled by PCR with digoxigenin-11-dUTP (Roche, Mannheim, Germany). The Fluorescence *in situ* hybridization was performed according to Cabral-de-Mello and Marec (2021) and probes labeled with biotin were detected with streptavidin Alexa-Fluor-488 conjugate (Invitrogen, CA, USA), while digoxigenin labeled probe was detected using anti-digoxigenin-rhodamine (Roche). Chromosomes were stained with DAPI and the slides mounted using VECTASHIELD (Vector, CA, USA). The results were analyzed using an Olympus BX61 epifluorescence microscope and photographs were recorded using a DP70 cooled digital camera. The images were merged and optimized for brightness and contrast using Adobe Photoshop CS2 software.

Results

A total of 43 families of satDNAs with "high confidence" according to TAREAN in at list one of the *A. pisum* populations were identified. The BLAST search with our collection did not find any significant match with other previously described repeats. The monomer length of the consensus sequences varied extensively from 19 bp to 3,661 bp and all satDNAs were A+T rich, mean of 66,3%, ranging from 51% (ApiSat12-194) to 78,2% (ApiSat39-1227). The molecular divergence for the satDNA families varied among the genomes from 0,21% to 20,53%. Considering each satellite, the highest range of divergence is the ApiSat42-562, presenting 0,39% to 20,53%. In comparison, the ApiSat16-718 presented the lowest range of divergence, 1,47% to 1,70%. The collection with the 43 satDNA families has an abundance of 0,83% of the genome on average, considering all populations, and it varied from 0,67% (*Pisum sativum*) to 1,03% (*Genista sagittalis*).

Among the 43 satDNAs, 41 were present in all populations of *A. pisum*, presenting a notable quantitative variation between them. Examples of the most pronounced quantitative variation are observed for the most abundant satDNAs, the satDNAs ApiSat01-173 that ranged from 0,1% (*Vicia cracca* biotype) to 0,19% (*Genista sagitallis* biotype) and the ApiSat02-3661 that ranged from 0,09% (*Genista tinctoria* biotype) to 0,27% (*Ononis spinosa* biotype). The ApiSat41-185 varied from 0,01% to 0,000005%, depending on the biotype. Interestingly some specific satDNAs revealed considerable abundance differences, such as the ApiSat43-174 that was only present in two populations, *M. officinalis* and *O. viciifolia* (Supplementary Table 2, Figure 1a). Part of satDNAs observed in *A. pisum* were also observed in the genomes of the sister species, being 27 in *A. craccivora, 33 in M. persicae* and 39 in *A. caraganae*, comprising to from about 0,05% to 1,2% of their genomes (Supplementary Table 2;

Figure 1b).

The analysis between more related populations, i.e. the six populations from the same biotype hosted in *Medicago sativa*, also revealed quantitative differences among the satDNAs, but as expected, to a lesser extent in comparison to all biotypes. The library abundance ranged from about 0,74% in *M. sativa RP* to 0,90% in *M. sativa LS*, mean 0,81% (Figure 2).

To check the putative patterns of more recent amplification among biotypes of specific monomers of satDNA families, that tend to show lower Kimura 2-parameter (higher homogeneity), we run the landscapes (abundance *versus* divergence) for each satDNA (Figure 3, Supplementary Figure 1). On the contrary, it is expected that satDNAs with old amplification are more variable (higher Kimura 2-parameter values). Some satDNAs showed a pattern of homogenization and more recent amplification only in one biotype or biotypes hosted on the same plant genus, like the ApiSat1-173 and ApiSat23-192 with a distinguished peak of amplification in the two biotypes of the *Genista* genus (Figure 3a,b). While other satDNAs showed similar patterns of homogenization and amplification among biotypes. These patterns included, for example, recent amplification for all biotypes, as ApiSat18-631 (Figure 3c); ancient amplification as for ApiSat17-19 (Figure 3d); variable amplification in distinct periods as for ApiSat38-147 (Figure 3e); and with two marcant amplifications, and more ancient and one more recent as for ApiSat03-334 (Figure 3f).

By the analysis on chromosome assemblies, we were able to identify 40 of the 43 satDNA families, the families ApiSat17-19, ApiSat42-562 and ApiSat43-174 were not identified (Supplementary Table 3), putatively because during the assembly some of the repeats could be excluded. Due to the elimination of some repeated copies, our analysis is probably an underestimation of satDNA abundances per chromosome, but as our main focus is to compare the satDNAs between chromosomes this technical

limitation does not hamper our conclusions. Our computational genomic analysis revealed that the distinct satDNAs were widely distributed along the axes of chromosomes with accumulation of specific families in some chromosomes (Figure 4, Supplementary Table 3). The Chr-X presented the highest number of satDNA families, 38 families, followed by the Chr-A1 with 32 families, Chr-A2 with 27 families and Chr-A3 with 20 families, that is correlated with chromosome size (Figure 4a,b). Considering the proportion that the satDNAs represented per chromosome bearing in mind their amount of DNA we observed that Chr-X showed about 0,27% of satDNAs, Chr-A1 has about 0,09%, and Chr-A2 and Chr-3 has similar abundances of satDNAs, i.e. about 0,12%. (Supplementary Table 3).

For cytogenetic investigation, we selected five repeats for FISH analysis as follows: (i) ApiSat01-173, due to its high abundance; (ii) ApiSat02-3661, due to its high abundance and low divergence; (iii) ApiSat03-334, due to its enrichment on autosomes and low abundance on X chromosome; (iv) ApiSat17-19, due to its high divergence; and (v) ApiSat24-549, due to its enrichment on X chromosome and low abundance on autosomes (Supplementary Tables 2, 3). The FISH for satDNAs ApiSat01-173 and ApiSat02-3661 revealed a large cluster near to one termini of the X chromosome, close to the rDNA cluster, plus one subterminal minor cluster in the opposite chromosome termini (Figure 5a,b). The ApiSat03-334 was located exclusively as one large cluster on one subtermini of the X chromosome, also near to the rDNA cluster (Figure 5c). We did not observe signals for ApiSat17-19 (Figure 5d) and for ApiSat24-549 scattered small clusters in all chromosomes were noticed (Figure 5e).

Discussion

Satellite DNA in the *Acyrthosiphon pisum* biotypes

At least 15 sympatric *A pisum* biotypes were identified in Western Europe (Peccoud et

al, 2015), which can compose different hybridization scenarios and even reproductive isolation (Drès & Mallet, 2002). The *A. pisum* biotypes are associated preferentially or exclusively to evolutionarily related host-plants, and they form gene flow clusters (Peccoud et al, 2015). Only the host species *Vicia faba* was observed as a resource for several biotypes (Ferrari et al, 2008; Peccoud et al, 2014). In this context, our collection showed that almost all satellite DNA families are shared between the populations, with two exceptions, and the variation was remarkably quantitative, with peaks of differential amplification in most cases in the *Genista* genus cluster. In some cases such as ApiSat04-253 and ApiSat19-248, the quantity increased by more than 5 times, depending on the biotype. These variations are in accordance with the patterns of evolution of satDNAs that could have an impact on the architecture of the genome (Garrido-Ramos, 2017).

Even a delicate variation could have consequences in the population genotype and in the success of hybrids. Considering the predominance of parthenogenetic reproduction in the life cycle of *A. pisum* and the possibility of superclones (Jaquiéry et al., 2012; Martel et al., 2020; Serteyn et al. 2021), these eventual changes on satDNAs are more easily fixed in populations. An interesting case is the ApisSat43-174 family, present in the biotypes *O. viiifolia* and *M. officinalis*, that could have emerged from a transformation of the most common ApiSat01-173 family. A homology analysis from these two satDNAs revealed 40% sequence alignment distributed in blocks of parity and this can be explained by the mechanisms of emergence, transformation and propagation of satellite sequences indicated by Ruiz-Ruano et al (2016) and Lower et al. (2018). The ApiSat42-562 family was absent in *L. corniculatus*, *V. cracca* and *M. sativa LS* populations, and it had more than 15% divergence in other 9 populations. Contrasting, less than 5% divergence in *O. viciifolia* and *O. spinosa* species, and less than 1% of divergence in the *Genista* genus.

The absence of the consensus sequence in a biotype can be explained by (i) the abundance is very small that it cannot be detected by the methods employed, (ii) it was degenerated (loss) in that biotype, or (iii) the consensus sequence appeared *de novo* in a biotype that had hybridization with only a part of the analyzed genomes. In this case, it is possible that this family is undergoing a process of degeneration in most populations and being maintained in others, or that it has homogenized in a small minority and shared by hybridization with only a few populations.

In addition to our collection of satDNAs, it is likely that other different families were ignored by the RepeatExplorer2 due to the minimum abundance threshold > 0.01%, necessary for detection (Novák et al., 2017). The majority of the satDNA families presented, at an intraspecific level, a picture predicted by the library hypothesis (Fry & Salser, 1977), where families are shared between genetic clusters and there is an eventual quantitative variation, a hypothesis also observed in other groups in different taxonomic levels (Plohl et al., 2012). When we look at the panorama of this collection in other aphid species, the number of missing sequences follows evolutionary distance, with marked quantitative variation. The *Genista* genus as host-species represent a possible scenario of speciation (Peccoud et al., 2015), and the peculiarities of the library registered here support this hypothesis. Belyayev and colleagues (2020a) proposes the emergence of a unique novel satellite family and their absence in proximate species opposes the library hypothesis, and suggests a shared "library of mechanisms of origin". In the recent work by Camacho et al. (2021) a possible model of recursive cycles of amplification, degeneration and rejuvenation of satDNA families, resulting in mostly contingent events where concerted evolution occurs in the lineages. Future homology analysis of our satDNA collection and the prospect of the satDNA from proximate species could contribute to the understanding of these two proposals.

C-Banding, CMA_3, silver staining, DAPI, and FISH analysis (Bizarro et al.,

2000) describes two heterochromatic bands on the X-chromosome telomeres, one A+T rich band, while the other is CMA_3 responsive and this last portion coincides with the NOR region. In addition, there is a telomeric heterochromatic band in the second pair of autosomes. Similarly, our computational mapping shows satDNA sequences widely dispersed along the chromosomes, especially on the X chromosome, with distinguishable accumulation of satDNA sequences near the terminal regions of the X chromosome, and a region of accumulation of these sequences also near to one end on chromosome A2. On the X chromosome, one of these regions appears to be more truncated, which seems to be in agreement with the heterogeneity of the heterochromatic regions that was observed under microscopy. An enrichment of satDNAs on the X chromosome was also proved here by FISH analysis with demonstration of occurrence of satDNA clusters on this chromosome. Moreover, we also noticed scattered satDNAs, as observed by computational analysis, revealing organizational plasticity of these elements on *A. pisum* genome.

Mathers et al. (2021) presented a rich analysis showing a conservation of the X chromosome among aphids proximate to *A. pisum* due a recalcitrant nature to recombination of this chromosome and intense recombination of the autosomes linked to break points associated to transposable elements. It can explain the higher variety of satDNA families observed in the X chromosome. According to Navajas-Pérez et al. (2009), the lack of recombination difficulties homogenization by concerted evolution. About the dispersed sequences in the autosomes, Ruiz-Ruano et al. (2016) proposes that a satDNA born due novel duplications of a sequence by different mechanisms and get disseminated by events such transposition, and Fernandéz et al. (2020) found evidences that the *A. pisum* genome get through duplication events and neofunctionalization of paralogous genes. Combining these finds with Mathers et al. work (2021), it is possible to imagine that sequences passed through duplication events such as the paralogous

genes and got scattered in the autosomes by transposition and recombination events, or some these transposable elements turned into satDNA families, as observed in other organisms (Dias et al., 2014; Belyayev et al., 2020b). It requires further comparative research, involving different biotypes, to investigate in depth these hypotheses along with eventual function of satDNAs in this species with great plasticity in genome and morphology and understudied holocentric chromosomes. A possible next step is to compare using cytogenomic tools the satDNA library between divergent lineages, e.g. *Genista* genus, to investigate the maintenance of translocation of the satDNA families observed in a *M. sativa* lineage.

Conclusions

Altogether, our data reveals that the satDNAs evolve rapidly even at intraspecies level, mainly by quantitative changes as predicted by the Library Hypothesis. But also, the origin of new repeats and disappearance of satDNA families could occur before the complete reproductive isolation, i.e. speciation process, as we noticed this phenomena in *A. pisum*. The data also shows that the holocentric structure of the chromosomes could influence in the evolution of satDNAs families, hampering the emergence of large chromosomal clusters, and homogenization of repeats, due to the absence of specific structure (centromere) with accumulation of repeats. Although, there are exceptions, as the case of the X chromosome with accumulation of specific repeats.

References

Afgan E et al. 2016. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Res. 44. doi: 10.1093/nar/gkw343.

Bachmann L, Raab M, Sperlich D. 2009. Satellite DNA and speciation: A species specific satellite DNA of *Drosophila* guanchel. J. Zool. Syst. Evol. Res. 27. doi: 10.1111/j.1439-0469.1989.tb00333.x.

Belyayev A et al. 2020a. The major satellite DNA families of the diploid *Chenopodium album* aggregate species: Arguments for and against the "library hypothesis". PLoS ONE, 15. doi: 10.1371/journal.pone.0241206

Belyayev A et al. 2020b. Transposons and satellite DNA: on the origin of the major satellite DNA family in the *Chenopodium* genome. Mobile DNA. doi: 10.1186/s13100-020-00219-7.

Bizzaro D. et al. 2000. Chromosome analysis and molecular characterization of highly repeated DNAs in the aphid *Acyrthosiphon pisum* (Aphididae, Hemiptera). Genetica. doi: 10.1023/A:1004128314728.

Brisson JA, Stern DL. 2006. The pea aphid, *Acyrthosiphon pisum*: An emerging genomic model system for ecological, developmental and evolutionary studies. BioEssays. doi: 10.1002/bies.20436.

Cabral-de-Mello, D., Moura, R. & Martins, C., 2010. Chromosomal mapping of repetitive DNAs in the beetle *Dichotomius geminatus* provides the first evidence for an association of 5S rRNA and histone H3 genes in insects, and repetitive DNA similarity between the B chromosome and A complement. *Heredity* 104. https://doi.org/10.1038/hdy.2009.126

Cabral-de-Mello DC, Zrzavá M, Kubíčková S, Rendón P, Marec F. 2021. The Role of Satellite DNAs in Genome Architecture and Sex Chromosome Evolution in Crambidae Moths. Front. Genet. 12. doi: 10.3389/fgene.2021.661417.

Camacho JP et al. 2021. On the contingent nature of satellite DNA evolution. bioRxiv doi: 10.1101/2021.07.01.450629.

Charlesworth B, Sniegowski P, Stephan W. 1994. The evolutionary dynamics of repetitive DNA in eukaryotes. Nature. doi: 10.1038/371215a0.

Dias GB, Svartman M, Delprat A, Ruiz A, Kuhn GCS (2014). Tetris is a foldback transposon that provided the building blocks for an emerging satellite DNA of *Drosophila virilis*. Genome Biol Evol 6. doi:10.1093/gbe/evu108

Dover G. 2002. Molecular drive. Trends Genet. 18. doi: 10.1016/S0168-9525(02)02789-0.

Dover G. 1982. Molecular drive: a cohesive mode of species evolution. Nature. 299. doi: 10.1038/299111a0.

Dover GA. 1986. Molecular drive in multigene families: How biological novelties arise, spread and are assimilated. Trends Genet. doi: 10.1016/0168-9525(86)90211-8.

Drés M, Mallet J. 2002. DRÈS, Michele; MALLET, James. Host races in plant–feeding insects and their importance in sympatric speciation. Philosophical Transactions of the Royal Society of London. doi: 10.1098/rstb.2002.1059.

Elder JF, Turner BJ. 1995. Concerted evolution of repetitive DNA sequences in eukaryotes. Q. Rev. Biol. doi: 10.1086/419073.

Escudeiro A, Adega F, Robinson TJ, Heslop-Harrison JS, Chaves R. 2019. Conservation, divergence, and functions of centromeric satellite DNA families in the Bovidae. Biol. Evol. 11. doi: 10.1093/gbe/evz061.

Férnandez R et al. 2020. Selection following gene duplication shapes recent genome evolution in the pea aphid Acyrthosiphon pisum. Molecular biology and evolution. doi: doi.org/10.1093/molbev/msaa110.

Ferrari J, Via S, Godfray HCJ. 2008. Population differentiation and genetic variation in performance on eight hosts in the pea aphid complex. Evolution. doi: 10.1111/j.1558-5646.2008.00468.x.

Ferree PM, Prasad S. 2012. How Can satellite DNA divergence cause reproductive isolation? Let us count the chromosomal ways. Genet. Res. Int. 2012. doi: 10.1155/2012/430136.

Fry K, Salser W. 1977. Nucleotide sequences of HS-α satellite DNA from kangaroo rat dipodomys ordii and characterization of similar sequences in other rodents. Cell. 12. doi: 10.1016/0092-8674(77)90170-2.

Garrido-Ramos M. 2017. Satellite DNA: An evolving topic. Genes (Basel). 8. doi: 10.3390/genes8090230.

Gel B., Serra E. (2017). karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. Bioinformatics. doi:10.1093/bioinformatics/btx346

Gonzáles M et al. (2020). Genomic differentiation of Deschampsia antarctica and D. cespitosa (Poaceae) based on satellite DNA. Botanical Journal of the Linnean Society. doi.org/10.1093/botlinnean/boaa045

Jaquiéry J et al. 2012. Accelerated evolution of sex chromosomes in aphids, an X0 system. Molecular biology and evolution. doi: 10.1093/molbev/msr252.

Li Y, Park H, Smith TE, Moran NA, Singh N. 2019. Gene family evolution in the pea aphid based on chromosome-level genome assembly. Mol. Biol. Evol. doi: 10.1093/molbev/msz138.

López-Flores I, Garrido-Ramos MA. 2012. The repetitive DNA content of eukaryotic genomes. Genome Dyn. doi: 10.1159/000337118.

Louzada S et al. 2020. Decoding the role of satellite DNA in genome architecture and plasticity—an evolutionary and clinical affair. Genes (Basel). doi: 10.3390/genes11010072.

Lower S et al. 2018. Satellite DNA evolution: old ideas, new approaches. Current opinion in genetics & development. doi:10.1016/j.gde.2018.03.003.

Malik HS, Henikoff S. 2001. Adaptive evolution of Cid, a centromere-specific histone in Drosophila. Genetics. doi: 10.1093/genetics/157.3.1293.

Martel S et al. 2020. Latitudinal trend in the reproductive mode of the pea aphid Acyrthosiphon pisum invading a wide climatic range. Ecology and Evolution. doi: 10.1002/ece3.6536.

Mathers T et al. 2021.Chromosome-scale genome assemblies of aphids reveal extensively rearranged autosomes and long-term conservation of the X chromosome. Molecular biology and evolution. doi: 10.1093/molbev/msaa246.

Meštrović N, Plohl M, Mravinac B, Ugarković D. 1998. Evolution of satellite DNAs from the genus Palorus - Experimental evidence for the 'library' hypothesis. Mol. Biol. Evol. doi: 10.1093/oxfordjournals.molbev.a026005.

Navajas-Pérez et al. 2009. Effect of location, organization, and repeat-copy number in satellite-DNA evolution. Molecular Genetics and Genomics. doi:10.1007/s00438-009-0472-4.

Novák P et al. 2017. TAREAN: A computational tool for identification and characterization of satellite DNA from unassembled short reads. Nucleic Acids Res. doi: 10.1093/nar/gkx257.

Palacios-Gimenez OM et al. 2020. Eight million years of satellite DNA evolution in grasshoppers of the genus *Schistocerca* illuminate the ins and outs of the library hypothesis. Genome Biol. Evol. doi: 10.1093/gbe/evaa018.

Peccoud J et al. 2008. Host range expansion of an introduced insect pest through multiple colonizations of specialized clones. Molecular Ecology. doi:10.1111/j.1365-294X.2008.03949.x .

Peccoud J et al. 2015. Genetic characterisation of new host-specialised biotypes and novel associations with bacterial symbionts in the pea aphid complex. Insect

Conservation and Diversity. doi: 10.1111/icad.12131

Pita S, Panzera F, Sánchez A, Palomeque T, Lorite P. 2017. Chromosome painting in triatomine insects reveals shared sequences between X chromosomes and autosomes. J. Med. Entomol. 54. doi: 10.1093/jme/tjw146.

Plohl M et al. 2010. Long-term conservation vs high sequence divergence: The case of an extraordinarily old satellite DNA in bivalve mollusks. Heredity (Edinb). doi: 10.1038/hdy.2009.141.

Plohl M, Luchetti A, Meštrović N, Mantovani B. 2008. Satellite DNAs between selfishness and functionality: Structure, genomics and evolution of tandem repeats in centromeric (hetero)chromatin. Gene. doi: 10.1016/j.gene.2007.11.013.

Ribeiro T et al. 2017. Centromeric and non-centromeric satellite DNA organisation differs in holocentric *Rhynchospora* species. Chromosoma. doi: 10.1007/s00412-016-0616-3.

Ruiz-Ruano FJ, López-León MD, Cabrero J, Camacho JPM. 2016. High-throughput analysis of the satellitome illuminates satellite DNA evolution. Sci. Rep. 6. doi: 10.1038/srep28333.

Salser W et al. 1976. Investigation of the organization of mammalian chromosomes at the DNA sequence level. Fed. Proc. 35. PMID: 1107072.

Serteyn L et al. 2021. Multi‐approach comparative study of the two most prevalent genotypes of pea aphid *Acyrthosiphon pisum* (Hemiptera: Aphididae) in Chile. Entomological Science. doi: 10.1111/ens.12436.

Shi J et al. 2010. Widespread gene conversion in centromere cores. PLoS Biol. 8. doi: 10.1371/journal.pbio.1000327.

Smalec BM, Heider TN, Flynn BL, O'Neill RJ. 2019. A centromere satellite concomitant with extensive karyotypic diversity across the *Peromyscus* genus defies predictions of molecular drive. Chromosom. Res. doi: 10.1007/s10577-019-

09605-1.

Sproul J., Khost D.E., Eickbush D., Negm S., Wei X., Wong I., Larracuente A., 2020. Dynamic Evolution of Euchromatic Satellites on the X Chromosome in *Drosophila melanogaster* and the *simulans* Clade, Molecular Biology and Evolution, Volume 37, Issue 8. doi.org/10.1093/molbev/msaa078.

Subirana JA, Albà MM, Messeguer X. 2015. High evolutionary turnover of satellite families in Caenorhabditis. BMC Evol. Biol. doi: 10.1186/s12862-015-0495-x.

Thakur J, Packiaraj J, Henikoff S. 2021. Sequence, chromatin and evolution of satellite DNA. Int. J. Mol. Sci. 22. doi: 10.3390/ijms22094309.

Ugarković Đ, Plohl M. 2002. Variation in satellite DNA profiles—causes and effects. EMBO J. 21. doi: 10.1093/emboj/cdf612.

Walsh JB. 1987. Sequence-dependent gene conversion: can duplicated genes diverge fast enough to escape conversion? Genetics. doi: 10.1093/genetics/117.3.543.

Supplementary Table 1.

Biotype or species	Accession number	Sequencer/strategy	Read length	GC content	Gbp	submitted by
Vicia cracca	SRX4038418	Illumina HiSeq 2000WGSPaired	100	29.4%	36.9	INRA
Onobrychis viciifolia	SRX4038417	Illumina HiSeq 2000WGSPaired	100	29.6%	41.4	INRA
Securigea varia	SRX4038416	Illumina HiSeq 2000WGSPaired	100	29.5%	39.0	INRA
Ononis spinosa	SRX4038415	Illumina HiSeq 2000WGSPaired	100	29.5%	43.3	INRA
Medicago lupulina	SRX4038414	Illumina HiSeq 2000WGSPaired	100	31.1%	41.8	INRA
Melilotus officinalis	SRX4038413	Illumina HiSeq 2000WGSPaired	100	29.4%	38.5	INRA
Lotus corniculatus	SRX4038412	Illumina HiSeq 2000WGSPaired	100	29.2%	45.2	INRA
Genista tinctoria	SRX4038411	Illumina HiSeq 2000WGSPaired	100	29.8%	42.7	INRA
Genista sagitallis	SRX4038410	Illumina HiSeq 2000WGSPaired	100	30.0%	37.4	INRA
Pisum sativum	SRX3185869	Illumina HiSeq 2000WGSPaired	100	28.8%	70.5	INRA
Medicago sativa Switzerland - SW	SRX4028530	Illumina HiSeq 2000WGSPaired	100	29.8%	28.1	INRA
Medicago sativa Ranspach - RP	SRX4028529	Illumina HiSeq 2000WGSPaired	100	30.4%	28.0	INRA
Medicago sativa Mirecourt - MR	SRX4028528	Illumina HiSeq 2000WGSPaired	100	29.9%	27.0	INRA
Medicago sativa Castelnaudary - CS	SRX4028527	Illumina HiSeq 2000WGSPaired	100	30.0%	26.6	INRA
Medicago sativa Gers - GE	SRX4028526	Illumina HiSeq 2000WGSPaired	100	29.7%	23.8	INRA
Medicago sativa Lusignan - LS	SRX4028525	Illumina HiSeq 2000WGSPaired	100	30.0%	25.5	INRA
*Acyrthosiphon caraganae**	SRX6674381	Illumina MiSeq WGS Paired	151	30.5%	43.7	Belarusian State University
*Myzus persicae**	SRX6674382	Illumina MiSeq WGS Paired	151	30.3%	52.0	Belarusian State University
*Aphis craccivora**	SRX2888378	Illumina HiSeq 2500 WGS Paired	149	29.7%	1.4	Embrapa

Supplementary Table 1. List of biotypes of *Acyrthosiphon pisum* and sister species with the accession number in the NCBI database, sequencing strategy, GC content and submission author used in this work.

Supplementary Table 2

SatDNA families	Length (bp)	A+T (%)	Average Abundance (%)	Abundance Range (%)	Average Divergence (%)	Divergence Range (%)
ApiSat01-173	173	52.3	0.1383	0.19 – 0.10	8.5	8.91 – 6.7
ApiSat02-3661	3661	56.2	0.1616	0.27 – 0.09	0.3	0.47 – 0.21
ApiSat03-334	334	71.5	0.0447	0.05 – 0.03	13.3	13.87 – 12.31
ApiSat04-253	253	69.3	0.0427	0.07 – 0.01	3.1	3.48 – 2.89
ApiSat05-1174	1174	65.8	0.0267	0.03 – 0.02	2.8	3.39 – 1.66
ApiSat06-1828	1828	60.5	0.0250	0.03 – 0.02	2.8	3.35 – 2.26
ApiSat07-315	315	65.1	0.0288	0.04 – 0.02	6.8	7.17 – 6.34
ApiSat08-1172	1172	65.5	0.0119	0.02 – 0.009	2.7	3.39 – 1.54
ApiSat09-786	786	71.8	0.0089	0.01 – 0.005	4.0	4.98 – 3.56
ApiSat10-929	929	63.2	0.0170	0.02 – 0.01	4.3	4.86 – 3.15
ApiSat11-316	316	53.4	0.0121	0.03 – 0.007	4.8	7.97 – 3.11
ApiSat12-194	194	51.0	0.0249	0.04 – 0.01	7.3	7.75 – 6.05
ApiSat13-668	668	63.4	0.0071	0.01 – 0.004	4.0	4.52 – 3.67
ApiSat14-203	203	59.6	0.0218	0.03 – 0.0003	4.0	10.97 – 2.79
ApiSat15-152	152	59.7	0.0145	0.02 – 0.006	7.3	9.43 – 5.79
ApiSat16-718	718	73.8	0.0095	0.013 – 0.006	1.6	1.70 – 1.47
ApiSat17-19	19	62.5	0.0123	0.02 – 0.008	18.8	19.35 – 17.76
ApiSat18-631	631	77.4	0.0107	0.016 – 0.006	2.5	3.01 – 2.15
ApiSat19-248	248	67.1	0.0127	0.023 – 0.003	3.1	6.24 – 2.30
ApiSat20-1634	1634	61.3	0.0101	0.018 – 0.005	2.3	2.79 – 1.04
ApiSat21-48	48	69.0	0.0183	0.027 – 0.007	5.6	5.97 – 5.36
ApiSat22-226	226	63.5	0.0038	0.007 – 0.002	3.1	5.08 – 2.41
ApiSat23-192	192	70.1	0.0231	0.09 – 0.006	7.0	7.52 – 6.50
ApiSat24-326	326	65.9	0.0104	0.017 – 0.005	5.7	8.07 – 5.06
ApiSat25-587	587	59.9	0.0099	0.014 – 0.006	4.2	5.56 – 3.23
ApiSat26-926	926	63.9	0.0097	0.014 – 0.004	4.4	9.28 – 2.60
ApiSat27-498	498	67.9	0.0060	0.009 – 0.004	3.7	6.99 – 2.48

ApiSat28-131	131	61.1	0.0079	0.010 – 0.006	9.0	10.14 – 7.86
ApiSat29-476	476	71.9	0.0051	0.018 – 0.002	2.3	3.06 – 1.21
ApiSat30-191	191	68.5	0.0197	0.031 – 0.005	6.6	7.15 – 6.06
ApiSat31-146	146	77.5	0.0148	0.022 – 0.004	6.9	12.73 – 4.08
ApiSat32-140	140	74.4	0.0095	0.023 – 0.004	5.7	7.48 – 3.26
ApiSat33-306	306	74.5	0.0069	0.013 – 0.003	3.4	3.96 – 2.82
ApiSat34-549	549	73.4	0.0073	0.011 – 0.003	4.6	5.38 – 4.34
ApiSat35-372	372	75.3	0.0051	0.009 – 0.002	1.9	1.18 – 2.69
ApiSat36-277	277	62.7	0.0012	0.0028 – 0.0005	4.8	5.41 – 4.57
ApiSat37-755	755	64.6	0.0091	0.062 – 0.001	9.7	12.55 – 1.0
ApiSat38-147	147	77.1	0.0036	0.005 – 0.001	7.6	11.21 – 5.81
ApiSat39-1227	1227	78.2	0.0090	0.014 – 0.001	8.0	15.97 – 5.22
ApiSat40-406	406	54.3	0.0039	0.020 – 0.001	12.4	19.45 – 4.11
ApiSat41-185	185	62.7	0.0016	0.01 – 0.000005	3.0	12.74 – 0.61
ApiSat42-562	562	74.4	0.0005	0.004 – 0	13.6	20.53 – 0.39
ApiSat43-174	174	69.6	0.0013	0.01 – 0	4.9	5.08 – 4.78

Supplementary Table 2. General genomic characteristics for the 43 satDNA families identified on the genome of *Acyrthosiphon pisum* by the analysis of short reads genomic libraries (Supplementary Table 1). Name of satDNA family, monomer size in base pairs, A+T content, mean and range of abundance and divergence among biotypes.

Supplementary Table 3.

satDNA families	Chr-A1		Chr-A2		Chr-A3		Chr-X	
	Divergence	Abundance	Divergence	Abundance	Divergence	Abundance	Divergence	Abundance
ApiSat21-48	0	0,00005623%	0	0,00000000%	0	0	4,84	0,00222415%
ApiSat08-1172	0	0,00000000%	2,71	0,00731711%	0	0	3,4	0,00072806%
ApiSat05-1174	13,31	0,00327514%	20,04	0,00335029%	0	0	17,45	0,00491834%
ApiSat39-1227	23,8	0,00407226%	22,22	0,00589919%	21,26	0,00338738%	22,69	0,00663474%
ApiSat28-131	11,38	0,00490100%	8,03	0,01383199%	10,87	0,00038031%	11,94	0,01019881%
ApiSat32-140	8,1	0,00385673%	11,77	0,00499323%	9,19	0,00107716%	12,77	0,00549701%
ApiSat31-146	24,21	0,00421517%	21,25	0,00540481%	23,69	0,00287478%	23,55	0,00736732%
ApiSat38-147	12,62	0,00034087%	0	0,00000000%	0	0	8,89	0,00009054%
ApiSat15-152	22,29	0,00400139%	18,55	0,00288100%	16,02	0,00122125%	11,05	0,01437476%
ApiSat20-1634	0	0,00000000%	0	0,00000000%	0	0,00000000%	4,01	0,01081898%
ApiSat01-173	11,61	0,00888951%	14,93	0,00869654%	18,59	0,00439367%	12,15	0,02453132%
ApiSat06-1828	23,16	0,00609638%	19,02	0,00745848%	24,94	0,00605901%	7,2	0,02766762%
ApiSat41-185	0	0,00000000%	0	0,00000000%	0	0,00000000%	26,42	0,00010034%
ApiSat30-191	1,72	0,00019152%	0,49	0,00393754%	4,62	0,00352675%	17,7	0,00319062%
ApiSat23-192	9,57	0,00011479%	19,07	0,00335615%	10,63	0,00577555%	3,16	0,00559735%
ApiSat12-194	24,1	0,00257232%	0	0,00000000%	0	0	27,86	0,00469803%
ApiSat14-203	19,5	0,00027176%	5,85	0,00007696%	16,37	0,00045590%	0	0
ApiSat22-226	13,52	0,00002518%	19,22	0,00010206%	0	0	2,01	0,00841300%
ApiSat19-248	26,99	0,00041408%	21,12	0,00385890%	0	0	20,68	0,00140330%
ApiSat04-253	2,99	0,00051774%	0	0,00000000%	0	0	10,8	0,00003772%
ApiSat36-277	0	0,00000000%	0	0,00000000%	0	0	9,29	0,00097476%
ApiSat33-306	22,72	0,00008551%	20,45	0,00005437%	0	0	4,19	0,00483761%
ApiSat07-315	12,54	0,00069872%	0	0,00000000%	6,5	0,00126849%	12,86	0,00140933%
ApiSat11-316	11,34	0,00976979%	33,09	0,00133175%	0	0	16,55	0,00517485%
ApiSat24-326	27,11	0,00016341%	0	0,00000000%	0	0	5,64	0,02525032%
ApiSat03-334	23,04	0,01563951%	24,74	0,04869261%	21,42	0,08642535%	18,26	0,00357766%
ApiSat02-3661	18,74	0,00005857%	0	0,00000000%	0	0	0	0
ApiSat75-372	19,15	0,00006677%	21,58	0,00015392%	0	0	2,16	0,00903770%
ApiSat40-406	26,6	0,00010835%	11,17	0,00034047%	27,31	0,00182125%	16,65	0,00953715%
ApiSat29-476	0	0,00000000%	21,5	0,00011126%	0	0	6,04	0,00955752%
ApiSat27-498	15,93	0,00019738%	0	0,00000000%	0	0	2,8	0,01102419%
ApiSat34-549	0	0,00000000%	30,3	0,00047097%	8,02	0,00059055%	25,86	0,00109850%
ApiSat25-587	27,42	0,00081761%	35,87	0,00117867%	20,88	0,00080787%	22,12	0,00502321%
ApiSat18-631	0	0,00000000%	13,83	0,00067257%	0	0	3,19	0,00075144%
ApiSat13-668	8,37	0,00371558%	31,69	0,00094779%	0	0	6,2	0,01006678%
ApiSat16-718	34,23	0,00024364%	0	0,00000000%	0	0,00000000%	26,58	0,00064959%
ApiSat37-755	0	0,00000000%	22,26	0,00017149%	19,51	0,00200550%	13,67	0,00934174%
ApiSat09-786	7,07	0,01125684%	0	0,00000000%	0	0	9,67	0,01718437%
ApiSat26-926	20,34	0,00295887%	20,35	0,00287598%	19,28	0,00242597%	15,55	0,00583123%
ApiSat10-929	10,91	0,00047967%	7,57	0,00794618%	0	0,00000000%	36,22	0,00028519%
Total abundance		0,09%		0,12816609%		0,12449672%		0,26910513%

Supplementary table 3. Satellite DNA families with their divergence and abundance percentages in each chromosome on the genome assembly of *Acyrthosiphon pisum*.

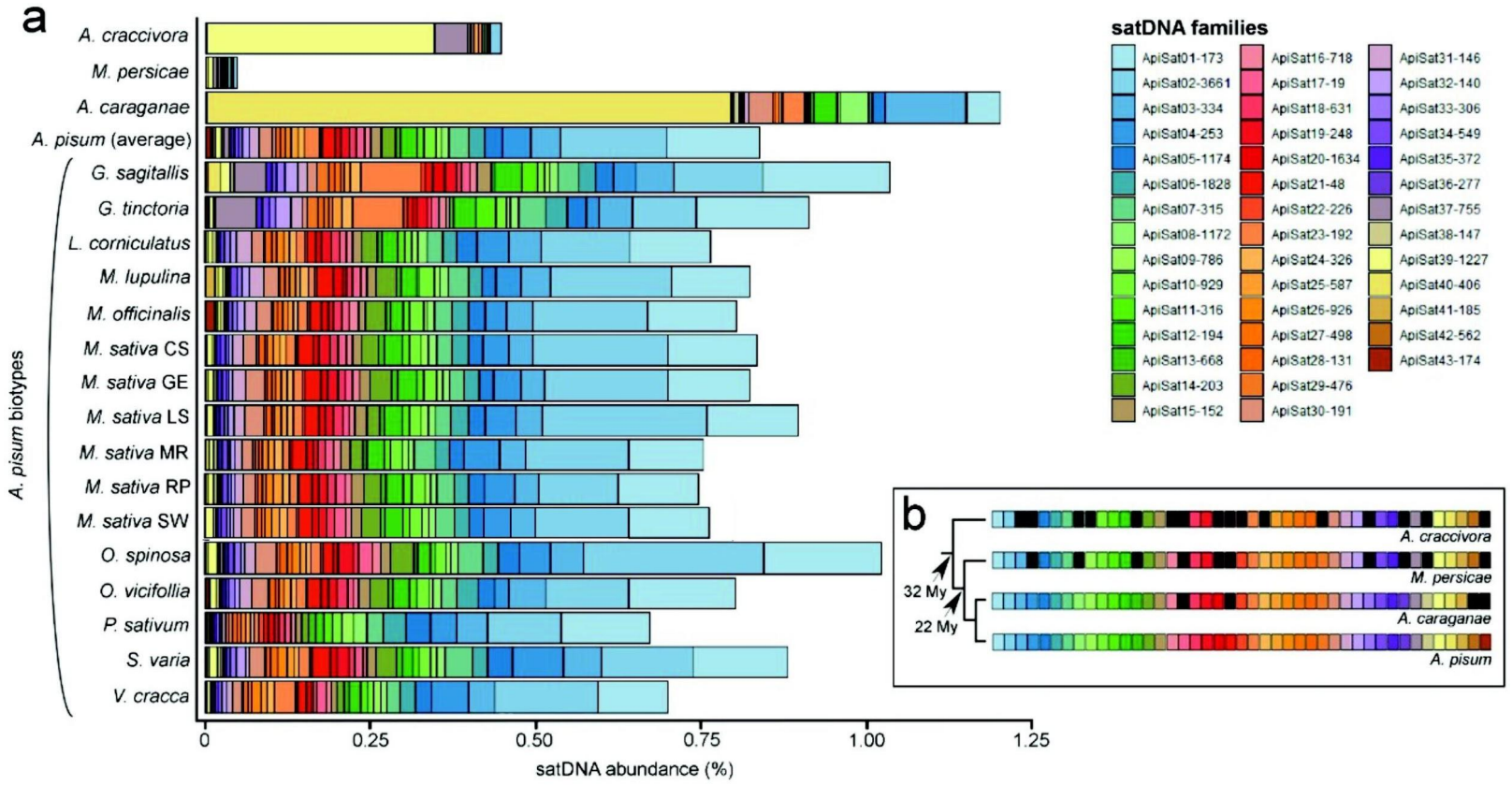


Figure 1. The satDNA library in *Acyrthosiphon pisum* and its satDNAs on sister species. (a) satDNA libraries in a stacked bar plots representing the satellite abundances in the sister species, the average abundance of the satDNA families in the *A. pisum* populations and in the *A. pisum* biotypes libraries. (b) Phylogenetic tree comparing *A. pisum* satDNAs with the other sister aphids analyzed. Absences of the satDNA families are represented by black squares. The phylogeny and time scale divergence is based on Methers et al. (2021). Each color corresponds to a family and each horizontal bar corresponds to the abundances stacked in every population.

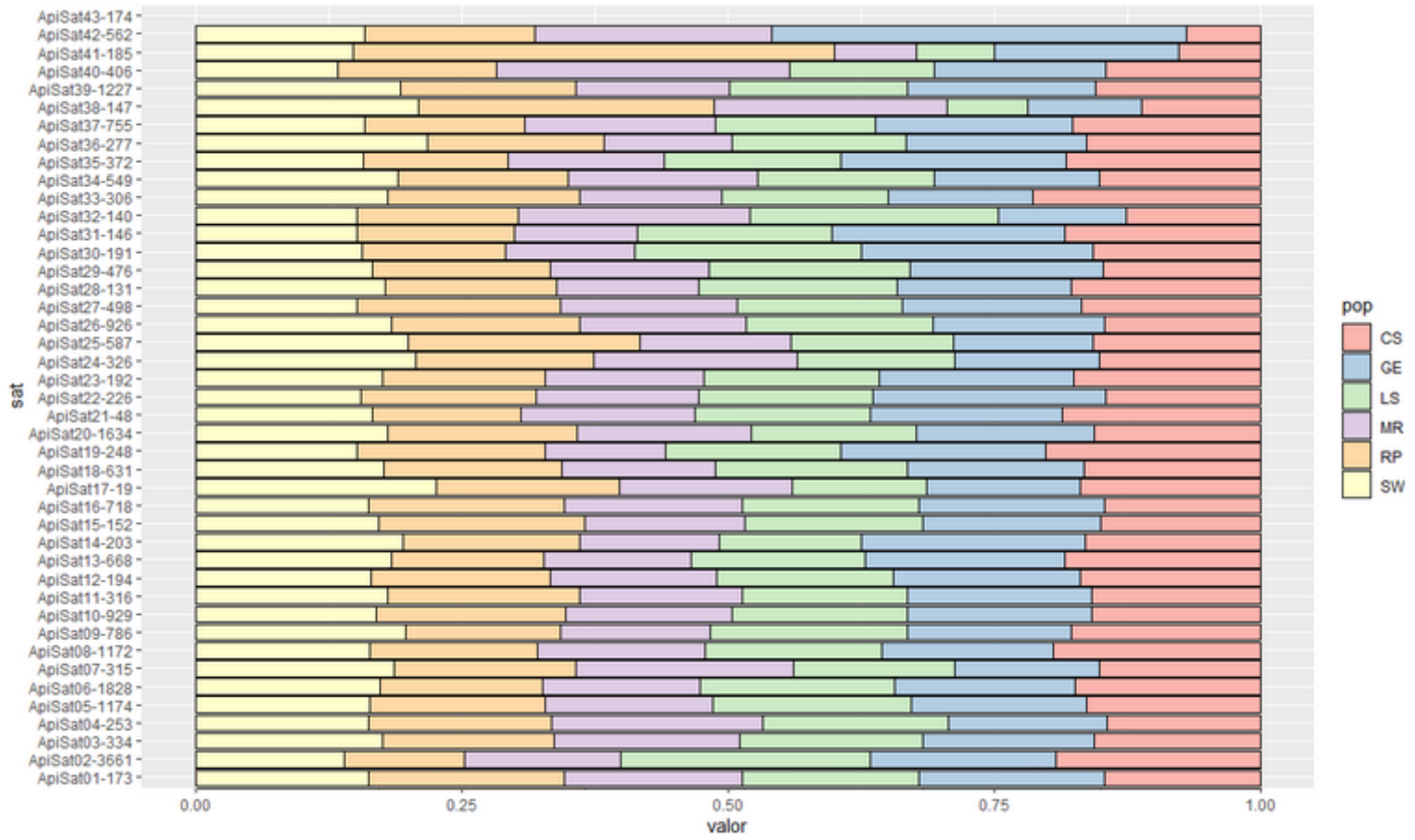


Figure 2. The satDNA library in *Acyrthosiphon pisum* biotypes hosted by plants from the same genus (*Medicago*). The proportions are shown in stacked bar plots for each satDNA and each color represents one region.

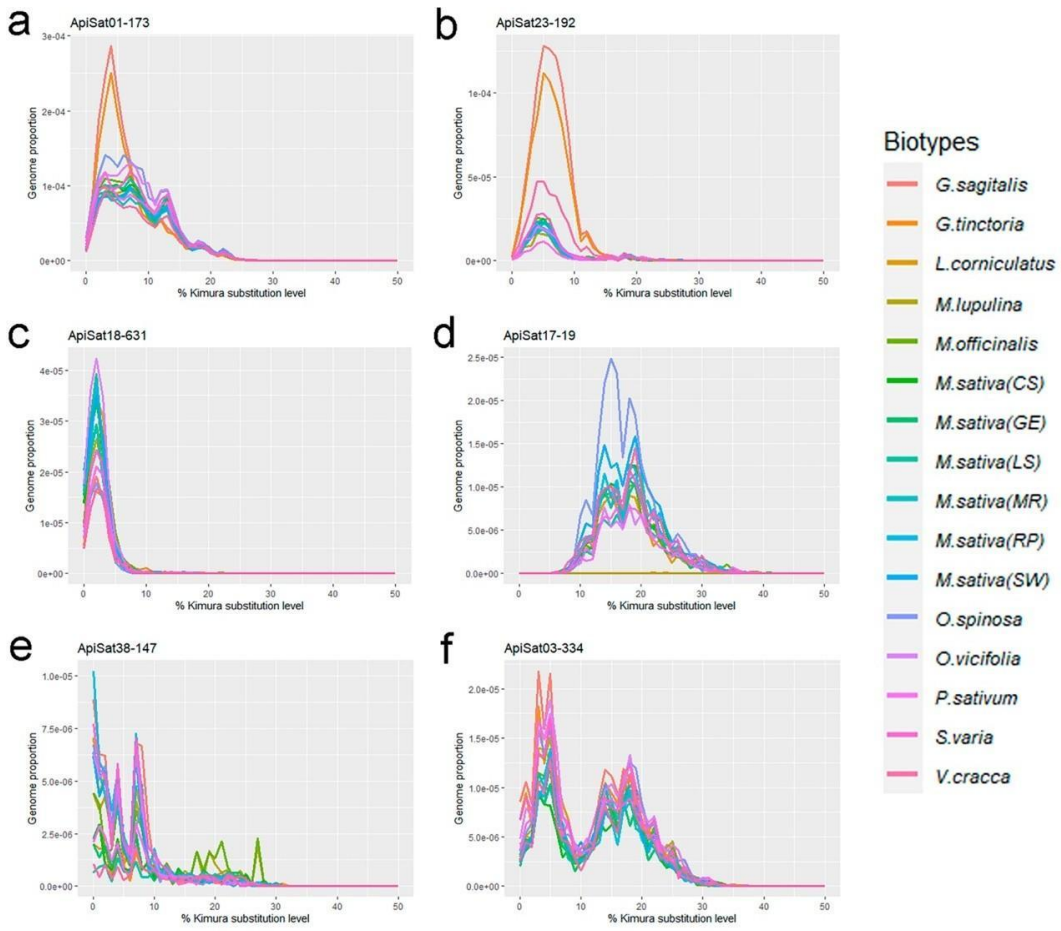


Figure 3. Landscapes with the abundances *versus* divergences of each satDNA family, colors correspond to host plant species. (a, b) ApiSat01-173 and ApiSat23-192 with recent amplification peak in *Genista* genus; (c) ApiSat18-631 with similar pattern of recent amplification in all biotypes; (d) ApiSat17-19 amplification peak with high K2P divergence (ancient amplification); (e) ApiSat38-147 with differential amplification peaks with distinct K2P; (f) ApiSat03-334 with two main peaks of amplification peaks in all biotypes.

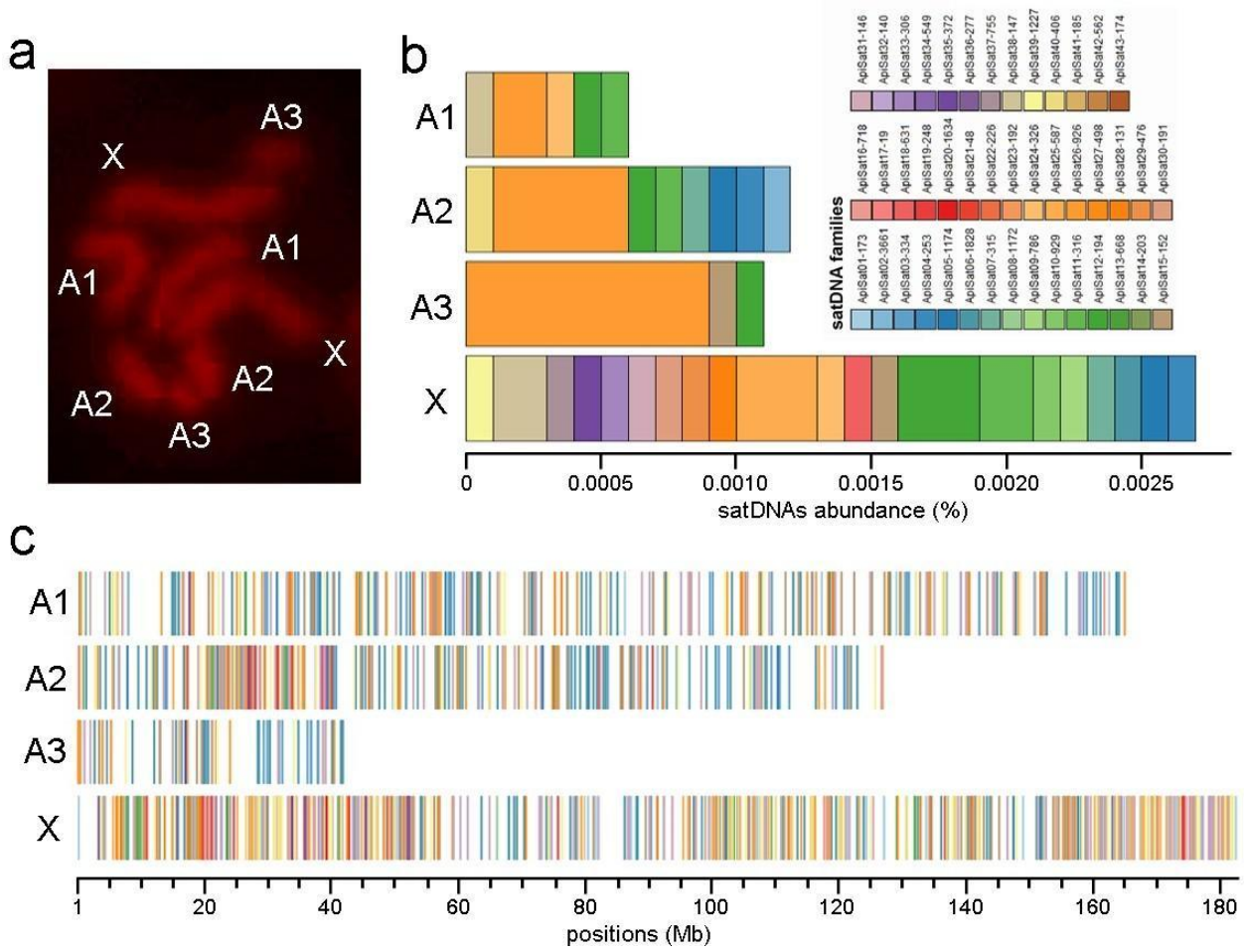


Figure 4. Chromosomal characteristics of satDNAs on *Acyrthosiphon pisum*. (a) *A. pisum* embryo chromosomes in stained with DAPI (pseudo colored in red) showing the diploid number of 2n = 8 and XX chromosomes (female); (b) stacked barplot representing the abundances of the satDNA families in each chromosome library, colors corresponding to each satDNA family; (c) representation of the *A. pisum* chromosome libraries in Megabases, each colored line correspond to a family of satDNA in a position. Note the dispersion of the satDNAs along the chromosomes with remarkable accumulation on the X chromosome.

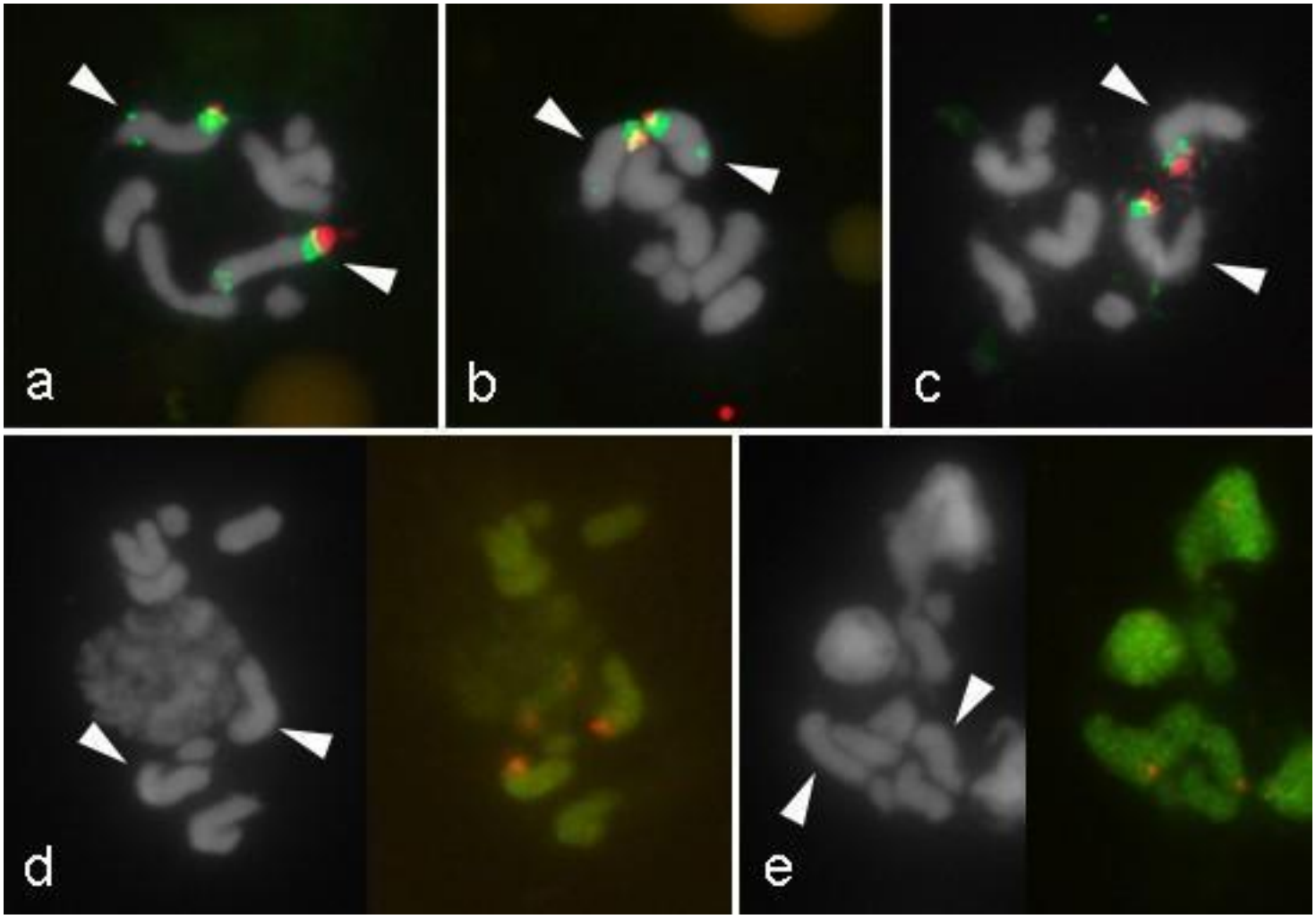
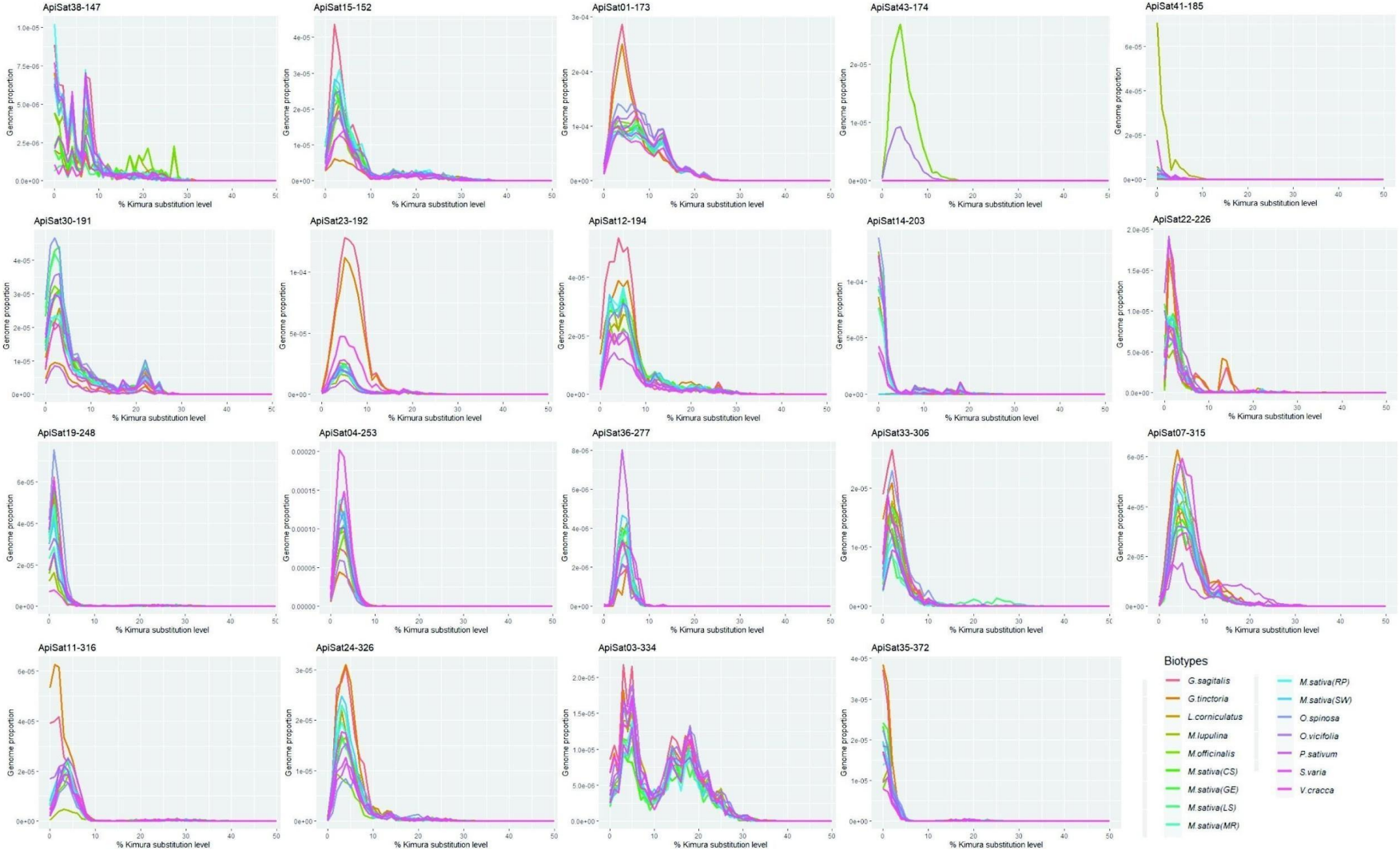
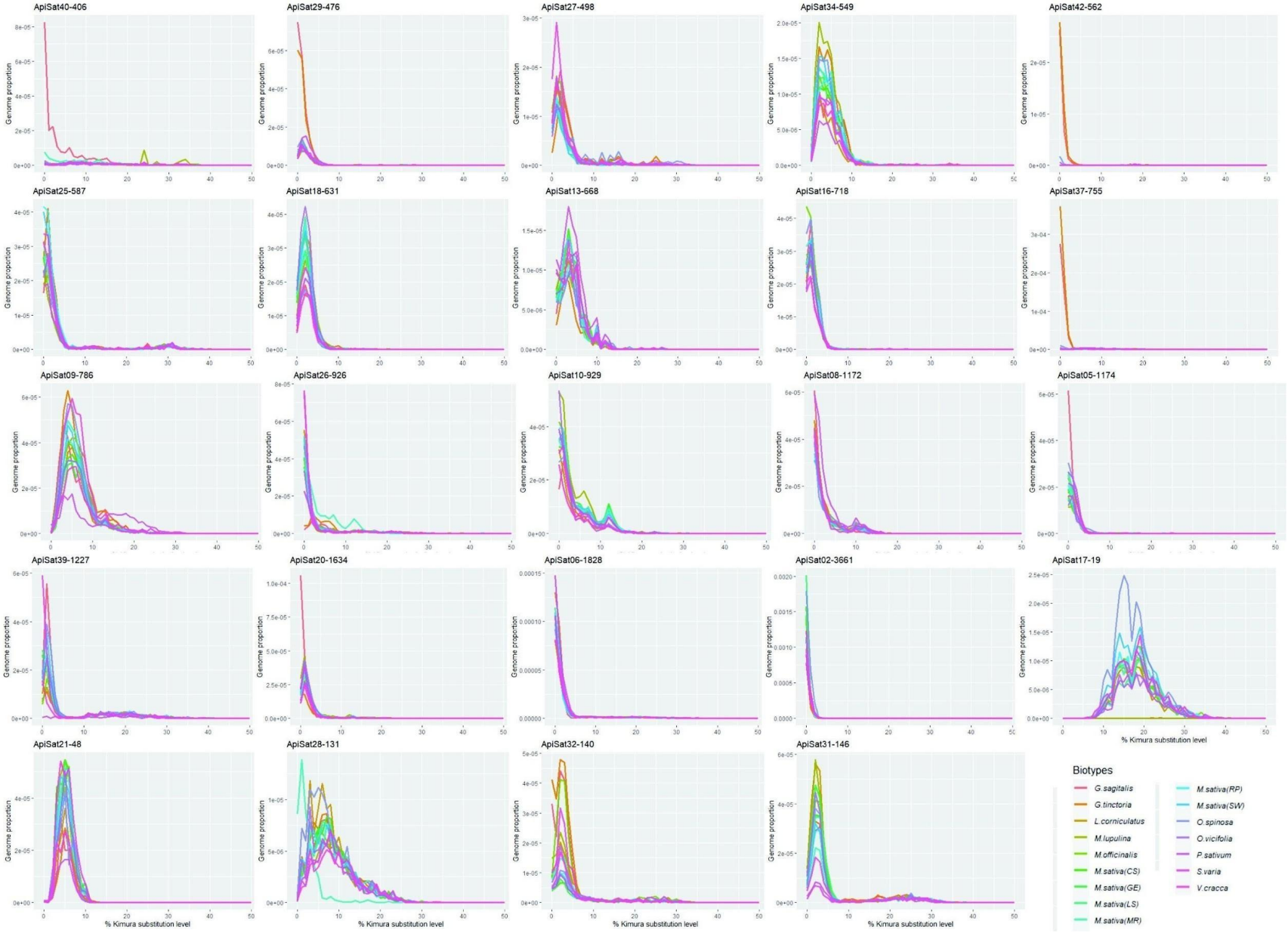


Figure 5. FISH mapping of five satDNAs (green) and 18S rDNA (red) in *Acyrthosiphon pisum* chromosomes from *Medicago sativa* biotype. (a) ApiSat01-173, (b) ApiSat02-3661, (c) ApiSat03-334, (d) ApiSat17-19, and (e) ApiSat24-549. The arrowheads show the X chromosomes in mitotic metaphase, identified by the rDNA cluster. In (d,e) on the left side are shown the chromosomes in DAPI, while on the right side the signals for satDNAs (green) and rDNA (red).

Supplementary figure 1. Landscapes representing abundance *versus* divergence for the 43 satDNA families in the 16 populations of *A. pisum*.




ApiSat40-406
ApiSat29-476
ApiSat27-498
ApiSat34-549
ApiSat42-562
ApiSat25-587
ApiSat18-631
ApiSat13-668
ApiSat16-718
ApiSat37-755
ApiSat09-786
ApiSat26-926
ApiSat10-929
ApiSat08-1172
ApiSat05-1174
ApiSat39-1227
ApiSat20-1634
ApiSat06-1828
ApiSat02-3661
ApiSat17-19
ApiSat21-48
ApiSat28-131
ApiSat32-140
ApiSat31-146
Genome proportion
% Kimura substitution level
Biotypes
G.sagitalis
G.tinctoria
L.corniculatus
M.lupulina
M.officinalis
M.sativa(CS)
M.sativa(GE)
M.sativa(LS)
M.sativa(MR)
M.sativa(RP)
M.sativa(SW)
O.spinosa
O.vicifolia
P.sativum
S.varia
V.cracca

4 References for main text

AFGAN E. et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. **Nucleic Acids Res. 44**, 2016. doi: 10.1093/nar/gkw343.

BACHMANN L, RAAB M, SPERLICH D. Satellite DNA and speciation: A species specific satellite DNA of *Drosophila* guanchel. **J. Zool. Syst. Evol. Res. 27**, 2009. doi: 10.1111/j.1439-0469.1989.tb00333.x.

BRISSON J.A., STERN D.L. The pea aphid, *Acyrthosiphon pisum*: An emerging genomic model system for ecological, developmental and evolutionary studies. **BioEssays**, 2006. doi: 10.1002/bies.20436.

CABRAL-DE-MELLO D.C., ZRZAVÁ M., KUBÍČKOVÁ S., RENDÓN P., MAREC F. The Role of Satellite DNAs in Genome Architecture and Sex Chromosome Evolution in Crambidae Moths. **Front. Genet. 12**, 2021. doi: 10.3389/fgene.2021.661417.

CHARLESWORTH B, SNIEGOWSKI P, STEPHAN W. The evolutionary dynamics of repetitive DNA in eukaryotes. **Nature**, 1994. doi: 10.1038/371215a0.

DOVER, G. Molecular drive. **Trends Genet**. 18:587–589, 2002. doi: 10.1016/S0168-9525(02)02789-0.

DOVER, G. Molecular drive: a cohesive mode of species evolution. **Nature**
299, 1982. doi: 10.1038/299111a0.

DOVER, G. Molecular drive in multigene families: How biological novelties arise, spread and are assimilated. **Trends Genet.**, 1986 doi: 10.1016/0168-9525(86)90211-8.

ELDER JF, TURNER BJ. Concerted evolution of repetitive DNA sequences in eukaryotes. **Q. Rev. Biol.**, 1995. doi: 10.1086/419073.

ESCUDEIRO A, ADEGA F, ROBINSON TJ, HESLOP-HARRISON JS, CHAVES R. Conservation, divergence, and functions of centromeric satellite DNA families in the Bovidae. **Biol. Evol. 11**, 2019. doi: 10.1093/gbe/evz061.

FERRARI J, VIA S, GODFRAY HCJ. Population differentiation and genetic variation in performance on eight hosts in the pea aphid complex. **Evolution**, 2008. doi: 10.1111/j.1558-5646.2008.00468.x.

FERREE PM, PRASAD S. How Can satellite DNA divergence cause reproductive isolation? Let us count the chromosomal ways. **Genet. Res. Int.**, 2012. doi: 10.1155/2012/430136.

FRY K, SALSER W. Nucleotide sequences of HS-α satellite DNA from kangaroo rat dipodomys ordii and characterization of similar sequences in other rodents. **Cell. 12**, 1977. doi: 10.1016/0092-8674(77)90170-2.

GARRIDO-RAMOS M. Satellite DNA: An evolving topic. **Genes (Basel). 8**, 2017. doi: 10.3390/genes8090230.

JAGANNATHAN M., YAMASHITA Y. Defective Satellite DNA Clustering into Chromocenters Underlies Hybrid Incompatibility in *Drosophila*, **Molecular Biology and Evolution, Volume 38, Issue 11**, 2021. doi.org/10.1093/molbev/msab221

LI Y, PARK H, SMITH TE, MORAN NA, SINGH N. Gene family evolution in the pea aphid based on chromosome-level genome assembly. **Mol. Biol. Evol.**, 2019. doi: 10.1093/molbev/msz138.

LÓPEZ-FLORES I, GARRIDO-RAMOS MA. The repetitive DNA content of eukaryotic genomes. **Genome Dyn.**, 2012. doi: 10.1159/000337118.

LOUZADA S. et al. Decoding the role of satellite DNA in genome architecture and plasticity—an evolutionary and clinical affair. **Genes (Basel).**, 2020. doi: 10.3390/genes11010072.

MALIK HS, HENIKOFF S. Adaptive evolution of Cid, a centromere-specific histone in Drosophila. **Genetics**, 2001. doi: 10.1093/genetics/157.3.1293.

MELTERS, D.P., PALIULIS, L.V., KORF, I.F. et al. Holocentric chromosomes: convergent

evolution, meiotic adaptations, and genomic analysis. **Chromosome Res. 20**, 2012. doi.org/10.1007/s10577-012-9292-1

MEŠTROVIĆ N, PLOHL M, MRAVINAC B, UGARKOVIĆ D. Evolution of satellite DNAs from the genus Palorus - Experimental evidence for the 'library' hypothesis. **Mol. Biol. Evol.**, 1998. doi: 10.1093/oxfordjournals.molbev.a026005.

NOVÁK P. et al. TAREAN: A computational tool for identification and characterization of satellite DNA from unassembled short reads. **Nucleic Acids Res.**, 2017. doi: 10.1093/nar/gkx257.

PALACIOS-GIMENEZ O.M. et al. Comparative analysis of morabine grasshopper genomes reveals highly abundant transposable elements and rapidly proliferating satellite DNA repeats. **BMC Biology**, 2020a. doi.org/10.1186/s12915-020-00925-x

PALACIOS-GIMENEZ O.M. et al. Eight million years of satellite DNA evolution in grasshoppers of the genus *Schistocerca* illuminate the ins and outs of the library hypothesis. **Genome Biol. Evol.**, 2020b. doi: 10.1093/gbe/evaa018.

PITA S, PANZERA F, SÁNCHEZ A, PALOMEQUE T, LORITE P. Chromosome painting in triatomine insects reveals shared sequences between X chromosomes and autosomes. **J. Med. Entomol. 54**, 2017. doi: 10.1093/jme/tjw146.

PLOHL M. et al. Long-term conservation vs high sequence divergence: The case of an extraordinarily old satellite DNA in bivalve mollusks. **Heredity (Edinb)**, 2010. doi: 10.1038/hdy.2009.141.

PLOHL M, LUCHETTI A, MEŠTROVIĆ N, MANTOVANI B. Satellite DNAs between selfishness and functionality: Structure, genomics and evolution of tandem repeats in centromeric (hetero)chromatin. **Gene**, 2008. doi: 10.1016/j.gene.2007.11.013.

RIBEIRO T. et al. Centromeric and non-centromeric satellite DNA organisation differs in holocentric *Rhynchospora* species. **Chromosoma**, 2017. doi:10.1007/s00412- 016-0616-3.

RUIZ-RUANO FJ, LÓPEZ-LEÓN MD, CABRERO J, CAMACHO JPM. High-throughput

analysis of the satellitome illuminates satellite DNA evolution. **Sci. Rep. 6**, 2016. doi:10.1038/srep28333.

SALSER W. et al. Investigation of the organization of mammalian chromosomes at the DNA sequence level. **Fed. Proc. 35**, 1976. PMID: 1107072.

SHI J. et al. Widespread gene conversion in centromere cores. **PLoS Biol. 8**, 2010. doi: 10.1371/journal.pbio.1000327.

SMALEC BM, HEIDER TN, FLYNN BL, O'NEILL RJ. A centromere satellite concomitant with extensive karyotypic diversity across the *Peromyscus* genus defies predictions of molecular drive. **Chromosom. Res.**, 2019. doi: 10.1007/s10577-019-09605-1.

SPROUL J., KHOST D.E., EICKBUSH D., NEGM S., WEI X., WONG I., LARRACUENTE A. Dynamic Evolution of Euchromatic Satellites on the X Chromosome in *Drosophila melanogaster* and the *simulans* clade. **Molecular Biology and Evolution, Volume 37, Issue 8**, 2020. doi.org/10.1093/molbev/msaa078

SUBIRANA JA, ALBÀ MM, MESSEGUER X. High evolutionary turnover of satellite families in Caenorhabditis. **BMC Evol. Biol.**, 2015. doi: 10.1186/s12862-015-0495-x.

THAKUR J, PACKIARAJ J, HENIKOFF S. Sequence, chromatin and evolution of satellite DNA. **Int. J. Mol. Sci. 22**, 2021. doi: 10.3390/ijms22094309.

UGARKOVIĆ Đ, PLOHL M.. Variation in satellite DNA profiles—causes and effects. **EMBO J. 21**, 2002. doi: 10.1093/emboj/cdf612.

WALSH JB. Sequence-dependent gene conversion: can duplicated genes diverge fast enough to escape conversion? **Genetics**, 1987. doi: 10.1093/genetics/117.3.543.