Next Generation Sequencing, Isolation and Characterization of 14 Microsatellite Loci of *Canthon cyanellus* (Coleoptera: Scarabaeidae)

Luis R. Arce-Valdés

Instituto de Ecología

Rosa Ana Sánchez-Guillén

Instituto de Ecología

Janet Nolasco-Soto

Instituto de Ecología

Mario E. Favila (✉ mario.favila@inecol.mx)

Institute of Ecology: Instituto de Ecología <https://orcid.org/0000-0002-2737-9327>

Research Article

Keywords: *Canthon cyanellus*, microsatellites, isolation, Mexico

Posted Date: May 10th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-489344/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background

We used Illumina paired-end-sequencing to isolate and characterize microsatellites of *Canthon cyanellus*, a Neotropical roller dung beetle encompassing several lineages within its distribution range.

Methods and Results

We examined *C. cyanellus* specimens collected at eight different localities in Mexico (two or three specimens per locality). We initially made amplification tests with 16 loci, but two of these showed amplification problems; thus, we evaluated 14 loci. The 14 microsatellites were polymorphic, with 2–16 alleles each. The expected and observed heterozygosity ranged from 0.11 to 0.76 and from 0.20 to 0.78, respectively.

Conclusions

These microsatellites will help to assess structure at the population and lineage levels, identify zones of potential hybridization between lineages, and draw a more precise geographic delimitation of *Canthon cyanellus* lineages.

Introduction

Canthon cyanellus Le Conte, 1859 is a Neotropical carrion roller beetle in the Subfamily Scarabaeinae; its distribution range stretches from the north of Brazil to Texas [1, 2]. The cuticular color of *Canthon cyanellus* varies widely across its range, from predominantly reddish-brown mostly in northern South America and Central America to entirely green or even blue in tropical regions of Mexico and Texas. The cuticular color variation allowed Halfpter [3] to establish three subspecies that can be distinguished mainly by differences in body color. *Canthon cyanellus cyanellus* LeConte, 1859 is distributed in Mexico along the Pacific and Gulf of Mexico slopes to Texas; it is mostly green, but green- or blue-colored specimens can be found in the northern plains of the Gulf of Mexico [3]. *Canthon cyanellus sallei* is distributed from Guatemala to Peru [4]; its dorsal surface is predominantly reddish-brown with greenish shades, elytra reddish-brown with greenish-black markings, and disc of the pronotum light colored, mainly reddish or yellowish-brown. *Canthon cyanellus violetae* is endemic to Chiapas in southern Mexico, with pronotum and pygidium largely reddish-brown [3].

The reproductive behavior, chemical communication, and ecology of *C. c. cyanellus* have been extensively studied [5–10, 10–19, 19–21]. Ortíz-Domínguez et al. [17] carried out experimental interpopulation crosses of *C. c. cyanellus* from the Gulf of Mexico coastal plains. These authors analyzed the composition of cuticular hydrocarbons used in sexual recognition and chemical communication [22]. Male-female interactions between individuals from populations geographically separated by more than 600 km were mainly aggressive. They fought for a food ball instead of rolling it cooperatively, as occurs between male-female pairs from the same population. Crosses between males and females from distant populations showed low fecundity and fertility, suggesting that *C. cyanellus* is undergoing an incipient reproductive isolation process.

Phylogeographic studies based on one nuclear (ITS2) and two mitochondrial (COI and 16S) loci resolved the *Canthon cyanellus* genealogy into seven clades from tropical regions of Mexico and one clade from Colombia [23, 24]. Five of such clades were well supported while the remaining ones were not. The geographic distribution of haplotypes suggests a limited historical gene flow and recent dispersal events between some clades [23, 24]. Nolasco-Soto et al. [24] used COI to estimate genetic distances between all the clades inferred for the group “cyanellus”. These authors found that the interclade divergence value surpassed the 2.2% threshold used to recognize separate insect species, according to Ratnasingham and Hebert [25]. Only two clades (SPS and SGM) were below such threshold (2.16%). However, the morphological variation of the aedeagus in the *C. cyanellus* populations studied by Nolasco et al. [24] did not coincide with the (lineages) clades; therefore, these could not be separated into different species. Studies by Nolasco-Soto et al. [23, 24] indicate that the phylogeographic history of *Canthon cyanellus* in Mexico is complex.

We used Illumina sequencing to isolate and characterize microsatellite loci of the *Canthon cyanellus* complex. Microsatellite markers (short DNA fragments, 100–500 bp) are polymorphic codominant markers that consist of tandem repeats of simple sequence motifs with high mutation rates (10^{-6} to 10^{-2} mutations per site per generation) [26]. Characteristics of microsatellite markers can contribute to better understand the population structure of this species and quantify the gene flow between clades/populations.

Methods

Sample Collection and DNA Extraction

Specimens of *C. cyanellus* were collected at eight localities in the Pacific and the Gulf of Mexico slopes, the Sierra Madre Oriental, and the southeast and southwest of the State of Chiapas (see Table 1 and Fig. 1). The geographic coordinates and elevation (masl) were recorded at each location. Twenty to thirty baited pitfall traps were placed at each location and left for 48 hours; traps were filled with damp soil and fish or squid as bait. All the beetles collected were transported alive to the laboratory, where they were subsequently frozen at -70°C. Both hind legs of each specimen were excised and ground to extract DNA following the DNeasy Blood & Tissue kit (QIAGEN) protocol.

Next-Generation Sequencing, De Novo Contig Assembly, and Selection of Microsatellites

Three samples of genomic DNA of *Canthon cyanellus* were used to construct the Illumina library following the Illumina Truseq DNA PCR-Free kit sequencing library protocol (Illumina Inc. USA). The library was sequenced using a 250-bp paired-end run on a MiSeq platform (Illumina Inc. USA). The quality of raw reads was checked with the software FastQC v0.11.3 [27] and filtered with cutadapt v1.8.3 [28]. Quality filtering consisted of clipping adapters, removing

sequences shorter than 20 nt, and trimming the 3' end to a threshold of $-q = 20$. The software SPAdes 3.10.1 [29] was used for *de novo* assembly and QDD 3.1.2 pipeline [30] for the detection, characterization, and primer design from the variable number tandem repeat (VNTR) loci; both programs were run with their default parameters.

The following criteria (based on Megléc et al. [30]) were used to select primer pairs for testing from the set of primer design sequences detected in the "Tuxpan" and "Los Tuxtlas" samples: (i) select only one primer pair per locus, (ii) the sequence should contain only pure microsatellites in the target region with the motifs AAT or AAG; (iii) the primer alignment score of the amplified sequence should be 2 or 3; (iv) markers should be selected from different ranges (at least 20 bp between each other) for PCR product length to facilitate multiplexing; (v) right and left primer lengths should be ≤ 60 bp.

Validation and Genotyping of Microsatellites

We performed a PCR to confirm the amplification and polymorphism of the loci selected. PCR was carried out with the universal fluorescent labeling method [31] on 18 *C. cyanellus* samples from eight populations (Table 1). PCR amplifications were made on a total reaction volume of 15 μ L using the GoTaq® Flexi DNA Polymerase PCR kit (Promega Corporation, USA). Each reaction contained 1X Green GoTaq® Flexi Buffer, 1.5 mM MgCl₂, 200 μ M dNTPs, 0.05 u/ μ L GoTaq® Polymerase, 0.5 μ M of reverse primer, 0.25 μ M of modified forward primer including the M13 sequence tail at the 5' end, 0.25 μ M of a labeled M13 forward primer with either FAM, HEX, or NED fluorescent probes, 1 μ L of genomic DNA (15-25 ng), and ddH₂O to 15 μ L. PCR reactions were standardized to the following thermal cycling conditions: 94°C for 3 minutes, followed by 30 cycles of 94°C for 1 minute, 50°C for 1 minute, and 70°C for 2 minutes; and a final extension of 72°C for 3 minutes. Automated fragment analysis of PCR products was conducted using gel capillary electrophoresis by MacroGen (MacroGen Inc., South Korea). Allele sizes were determined with the software GeneMarker 2.6.3 (SoftGenetics, USA), and binning was performed with the Autobin [32] Excel® macro.

Polymorphism Analyses

The total number of alleles (A), effective number of alleles (A_E), observed (H_O) and expected heterozygosity (H_E), inbreeding index (F_{IS}), and number of private alleles per locality were estimated using the macro GenAlEx 6.503 [33]. Finally, we calculated the polymorphism information content (PIC) and the probability of identity (P_{ID}) for each locus using the software PowerMarker 3.25 [34] and Gimlet 1.3.3 [35], respectively.

Results

Genome Sequencing and Characterization of Microsatellite Loci

Between six and eight million paired-end reads were generated per sample (Table 2). Most reads had the expected sequence length (250 bp) and were of sufficiently high quality to pass the filtering; only ca. 1% of the reads of each sample were filtered out.

Between 90,000 and 115,000 contigs were successfully assembled per sample using the accepted reads. The average contig length ranged from 2,000 to 3,000 bp with an N50 around 7,000 bp.

A total of about 9,000 VNTR loci were found per sample using the contigs; about 8,000 of them were identified as perfect or pure microsatellites. As per the QDD3 glossary, we defined pure microsatellites as loci composed of a single, 2–6 bp motif, with no interruption, and at least five repetitions [30]. Some 135,000 primer pairs were generated per sample from the VNTR loci found and characterized.

Overall, the VNTRs of three *C. cyanellus* samples from each locality had a very similar genomic characterization (Fig. 2). Almost 90% of all the VNTR loci found met the definition of perfect microsatellites; about 9% were categorized as compound microsatellites and the rest as minisatellite loci (Fig. 2a). Compound microsatellites were defined as microsatellite loci followed by 3–4 tandem repetitions of a 2–6 bp motif separated by a distance equal or shorter than the longest of two tandem repetitions. Minisatellite loci were defined as two or more perfect or compound microsatellite loci in the same target region [30].

The frequency of the VNTR classes was inversely related to motif size (Fig. 2b). Dinucleotide VNTRs accounted for approximately 65% of all the loci, followed by trinucleotides with about 30%, tetranucleotides with 5%, and so on. The average number of reiteration units seen in VNTR loci was close to the minimum threshold (five) employed by the QDD3 pipeline. The mean number of repetitions was 8.9 for dinucleotides, 5.5 for trinucleotides, 5.3 for tetranucleotides, 5.3 for pentanucleotides, and 8.4 for hexanucleotides; however, a higher number of repetitions was found in loci with shorter motifs (Fig. 2c; Dinucleotides: median = 5, Q3 = 9, Max = 66; Trinucleotides: median = 5, Q3 = 6, Max = 25; Tetranucleotides: median = 5, Q3 = 5, Max = 16).

We also found differences in the frequency of the various motifs in the VNTR loci. AT was the most common motif in dinucleotide VNTRs, and CG was the rarest (Fig. 2d); AAT and AGC were the most and least common motifs, respectively, in trinucleotide loci (Fig. 2e); AAAT was the most common motif in tetranucleotide loci, while ACAG and AACC were the rarest, with only one locus each.

Validation of Microsatellite Loci

Fourteen of the 16 primer pairs synthesized were successfully amplified and yielded single PCR products of the expected size; PCR amplification was not achieved in one locus and reading issues appeared in another one (SupplementaryMaterial2). The amplification success rate of the 14 markers was 95.2%, with only 12 missing data after amplifying 18 *C. cyanellus* samples. Most of these loci amplified fragments with lengths differing between them in three nucleotides. This result was expected, as primers were designed only on trinucleotide-perfect microsatellites; this facilitated the assignment of raw fragment

lengths — often referred to as binning — into allele classes [36]. However, Ccy09 and Ccy12 showed a mostly continuous distribution of raw fragment sizes with differences of less than three nucleotides. Based on this, we decided to bin their alleles into the mononucleotide allele class (Fig. 3).

Polymorphism Analyses

The number of alleles in polymorphic loci ranged from 2 to 16, and the observed heterozygosity from 0.11 to 0.82 (Table 3). Ten loci showed moderate polymorphism (4–6 alleles), four were marginally polymorphic (2–3 alleles), and only two were highly polymorphic (9 and 16 alleles). The effective number of alleles (A_E) was about half the number of observed alleles, suggesting the presence of a few dominant alleles in the loci.

Eight loci showed high (>0.20) absolute F_{IS} values, indicating large differences between their expected (H_E) and observed (H_O) heterozygosity. These loci displayed positive F_{IS} values, indicating a deficit of heterozygote samples vs. expected under HW equilibrium. Polymorphism information content (PIC) ranged from 0.190 to 0.889, with a mean of 0.494.

The moderate genetic diversity observed in these markers conferred a relatively high Probability of Identity (P_{ID}). This parameter measures the probability of two unrelated individuals having the same genotype [37]. P_{ID} values ranged from 60.1% (Ccy11) to 0.1% (Ccy12). However, when P_{ID} was computed over the full set of 14 markers using the product rule, it reached a value of 1.299×10^{-12} in unrelated individuals and 1.563×10^{-4} in siblings.

Since we examined only a few (2–3) samples per locality, we measured the ability of these 14 markers to measure genetic structure by identifying the number of private alleles of each locus. Eleven of the 14 loci possessed at least one private allele at one locality (Fig. 4). Ten of such 11 loci had few (1–3) private alleles, and locus Tuxt06 had eight private alleles across six different localities.

Discussion

Characterization of Microsatellite Loci

Perfect microsatellite loci were the most common type of VNTR in *Canthon cyanellus* genomes, as reported for the coconut rhinoceros beetle (*Oryctes rhinoceros*) [38]. However, a recent comparison of 29 beetle genomes showed that imperfect microsatellites were the most common short-sequence repeat type [39]. Our results and those from at least one other beetle genomic microsatellite characterization [40] are not consistent with this. A more reliable comparison between genomic characterizations of microsatellite loci requires the standardization of detection criteria and VNTR classes.

The most frequent VNTR class has been reported to be a highly heterogeneous feature of beetle genomes, particularly in Scarabaeidae species [39]. However, after excluding mononucleotide microsatellites (as we did in our analyses), dinucleotide and trinucleotide seem to be the most frequent microsatellite motif sizes [39, 41]. Our results are consistent with this observation as we found a trend for VNTR frequency to decrease with motif size. Our results are also consistent with the general pattern of short sequence repeats to favor poly(A) and poly(T) motifs over poly(G) and poly(C) in eukaryotic genomes [39, 42, 43].

Non-Tandem Repeat Polymorphisms in Ccy09 and Ccy12

The raw allele reads of Ccy09 and Ccy12 were not consistent with the expected three-nucleotide difference in size; this made us bin these loci as mononucleotide microsatellites. These markers might show evidence of non-tandem repeat polymorphisms, such as SNPs or indel mutations, and thus would not be perfect microsatellites. Complex mutational processes that produce variations in allele length at microsatellite loci have been reported in eukaryote groups, including yeasts [44, 45], fish [46], and insects [40].

Imperfect microsatellites have pros and cons when used as molecular markers for DNA fingerprinting and population genetics compared to perfect microsatellites [47]. Additional alleles provide more information that is useful for paternity testing, mapping studies, and allow achieving a more satisfactory resolution of population structure [47, 48]. Ccy09 and Ccy12 showed higher genetic diversity and PIC values than the other loci. However, the full advantages of imperfect microsatellites can only be achieved by using DNA sequencing. Moreover, mutational models of imperfect microsatellites are more complex than those of perfect microsatellites [47].

Polymorphism Analyses

Overall, our markers displayed moderate levels of genetic diversity and a deficit of heterozygote samples compared to what would be expected under HW equilibrium. The latter was expected due to the small sample size used to evaluate the primers and the use of samples from multiple localities to sample as many alleles as possible. Since a potential hidden genetic structure might be causing the heterozygote deficit — often referred to as the Wahlund effect [49] —, we decided not to test for departure from HW equilibrium.

More accurate estimates of diversity for each marker could be obtained by increasing the genotyped sample size. However, with this small sample size, we estimated a P_{ID} of 1.299×10^{-12} for unrelated individuals and 1.563×10^{-4} for siblings over the full set of 14 markers. In other words, the probability of sampling two unrelated individuals with the same genotypes across the 14 markers is 1 in ca. 770 billion and 1 in 6,000 for sibling beetles. This highlights the highly informative power of the full set of 14 microsatellite markers. Finally, the robustness of these markers to measure the genetic structure in *Canthon cyanellus* populations and lineages should be evaluated using larger samples from each locality.

Conclusions

The set of 14 *Canthon cyanellus* microsatellites that were isolated and characterized provide a valuable genetic tool to assess structure at the population or lineage levels and analyse hybridization processes. These microsatellites will help to better understand the geographic distribution of the lineages in this species complex. Further studies including populations from Central and South America would contribute to set a sound taxonomic delimitation of these lineages.

Declarations

Funding

This study was funded by a research grant from Consejo Nacional de Ciencia y Tecnología (CONACyT) awarded to Rosa Ana Sánchez Guillén (CB 282922).

Conflicts of interest/Competing interests

The authors declare to have no conflicts of interest that are relevant to the content of this article.

Availability of data and material

The database of microsatellite genotypes created and analyzed in this study is included in the SupplementaryMaterial1.xlsx file that accompanies the published article. Additional data supporting the findings of this study are available, upon request, from the corresponding author.

Code availability

Not applicable.

Authors' contributions

All the authors contributed to the conception and design of this study. Samples were collected by Janet Nolasco-Soto. Wet lab techniques were performed by Janet Nolasco-Soto and Luis Rodrigo Arce-Valdés. Data analyses were carried out by Rosa Ana Sánchez-Guillén and Luis Rodrigo Arce-Valdés. The manuscript was written and commented equally by all the authors. All the authors read and approved the final version of the manuscript.

Ethics approval

Not applicable.

Consent to participate

Not applicable.

Consent for publication

Neither the article nor portions of it have been previously published elsewhere. The manuscript is not under consideration for publication in any other journal and will not be submitted elsewhere until the Molecular Biology Reports editorial process is completed. All authors consent to the publication of the manuscript in Molecular Biology Reports, should the article be accepted.

Acknowledgements

Sequencing was performed at the Genomics Core Facility of Servicio Central de Soporte a la Investigación Experimental (SCSIE), Universidad de Valencia. This study was funded by a research grant (CB 282922) from Consejo Nacional de Ciencia y Tecnología (CONACyT) awarded to Rosa Ana Sánchez Guillén. María Elena Sánchez-Salazar edited the English manuscript.

References

1. Robinson M (1948) A Review of the Species of *Canthon* Inhabiting the United States (Scarabaeidae: Coleoptera). Trans Am Entomol Soc 1890- 74:83–100
2. Solís-Blanco A, Kohlmann B (2002) El género *Canthon* (Coleoptera: Scarabaeidae) en Costa Rica. The genus *Canthon* (Coleoptera: Scarabaeidae) in Costa Rica. G Ital Entomol 10:1–68
3. Halffter-Salas G (1961) Monografía de las especies norteamericanas del género *Canthon* Hoffsg.(Coleoptera: Scarabaeidae). Univ Zulia 20:225–320
4. Blackwelder RE (1944) Checklist of the coleopterous insects of Mexico, Central America, the West Indies, and South America. US Government Printing Office
5. Bellés X, Favila M (1983) Protection chimique du nid chez *Canthon cyanellus cyanellus* LeConte [Col. Scarabaeidae]. Bull Société Entomol Fr 88:602–607
6. Cortez V, Favila ME, Verdú JR, Ortiz AJ (2012) Behavioral and antennal electrophysiological responses of a predator ant to the pygidial gland secretions of two species of Neotropical dung roller beetles. Chemoecology 22:29–38
7. Cortez V, Verdú JR, Ortiz AJ, et al (2015) Chemical diversity and potential biological functions of the pygidial gland secretions in two species of Neotropical dung roller beetles. Chemoecology 25:201–213. <https://doi.org/10.1007/s00049-015-0189-2>

8. Favila M, Chamorro-Florescano I (2008) Male reproductive status affects contest outcome during nidification in *Canthon cyanellus cyanellus* LeConte (Coleoptera: Scarabaeidae). *Behaviour* 145:1811–1821. <https://doi.org/10.1163/156853908786279637>
9. Favila M, Chamorro-Florescano I (2009) The reproductive status of both sexes affects the frequency of mating and the reproductive success of males in the ball roller beetle *Canthon cyanellus cyanellus* (Coleoptera: Scarabaeidae). *Behaviour* 146:1499–1512. <https://doi.org/10.1163/156853909X445560>
10. Chamorro-Florescano IA, Favila ME, Macías-Ordóñez R (2011) Ownership, size and reproductive status affect the outcome of food ball contests in a dung roller beetle: when do enemies share? *Evol Ecol* 25:277–289
11. Chamorro-Florescano IA, Favila ME (2016) Male Success in Intrasexual Contests Extends to the Level of Sperm Competition in a Species of Dung Roller Beetle. *Ethology* 122:53–60. <https://doi.org/10.1111/eth.12443>
12. Chamorro-Florescano IA, Favila ME, Macías-Ordóñez R (2017) Contests over reproductive resources in female roller beetles: Outcome predictors and sharing as an option. *PLOS ONE* 12:e0182931. <https://doi.org/10.1371/journal.pone.0182931>
13. Favila ME (2001) Historia de vida y comportamiento de un escarabajo necrófago: *Canthon cyanellus cyanellus* LeConte (Coleoptera: Scarabaeinae). *Folia Entomológica Mex* 40:245–278
14. Favila ME, Ortiz-Domínguez M, Chamorro-Florescano I, Cortez-Gallardo V (2012) Comunicación química y comportamiento reproductor de los escarabajos rodadores del estiércol (Scarabaeinae: Scarabaeini): aspectos ecológicos y evolutivos, y sus posibles aplicaciones [pp. 141–164]. *Temas Sel En Ecol Quím Insectos JC Rojas EA Malo Ed El Col Front Sur Tapachula Mex*
15. Favila ME, Díaz A (1996) *Canthon cyanellus cyanellus* LeConte (Coleoptera: Scarabaeidae) Makes a Nest in the Field with Several Brood Balls. *Coleopt Bull* 50:52–60
16. Favila ME, Nolasco J, Florescano IC, Equihua M (2005) Sperm competition and evidence of sperm fertilization patterns in the carrion ball-roller beetle *Canthon cyanellus cyanellus* LeConte (Scarabaeidae: Scarabaeinae). *Behav Ecol Sociobiol* 59:38. <https://doi.org/10.1007/s00265-005-0006-y>
17. Ortiz-Domínguez M, Favila ME, Mendoza-López MR, et al (2006) Epicuticular compounds and sexual recognition in the ball-roller scarab, *Canthon cyanellus cyanellus*. *Entomol Exp Appl* 119:23–27. <https://doi.org/10.1111/j.1570-7458.2006.00388.x>
18. Ortíz-domínguez M, Favila ME, Mendoza-lópez MR (2006) Mate Recognition Differences Among Allopatric Populations of the Scarab *Canthon cyanellus cyanellus* (Coleoptera: Scarabaeidae). *Ann Entomol Soc Am* 99:1248–1256. [https://doi.org/10.1603/0013-8746\(2006\)99\[1248:MRDAAP\]2.0.CO;2](https://doi.org/10.1603/0013-8746(2006)99[1248:MRDAAP]2.0.CO;2)
19. Ortiz-Domínguez M, Favila-Castillo ME, González D, Zuñiga D (2010) Genetic differences in populations of the ball roller scarab *Canthon cyanellus cyanellus* (Coleoptera: Scarabaeidae): a preliminary analysis. *Elytron* 24:99–106
20. Arellano L, León-Cortés JL, Ovaskainen O (2008) Patterns of abundance and movement in relation to landscape structure: a study of a common scarab (*Canthon cyanellus cyanellus*) in Southern Mexico. *Landsc Ecol* 23:69–78. <https://doi.org/10.1007/s10980-007-9165-8>
21. Portela Salomão R, González-Tokman D, Dáttilo W, et al (2018) Landscape structure and composition define the body condition of dung beetles (Coleoptera: Scarabaeinae) in a fragmented tropical rainforest. *Ecol Indic* 88:144–151. <https://doi.org/10.1016/j.ecolind.2018.01.033>
22. Chung H, Carroll SB (2015) Wax, sex and the origin of species: Dual roles of insect cuticular hydrocarbons in adaptation and mating. *BioEssays* 37:822–830. <https://doi.org/10.1002/bies.201500014>
23. Nolasco-Soto J, González-Astorga J, Espinosa de los Monteros A, et al (2017) Phylogeographic structure of *Canthon cyanellus* (Coleoptera: Scarabaeidae), a Neotropical dung beetle in the Mexican Transition Zone: Insights on its origin and the impacts of Pleistocene climatic fluctuations on population dynamics. *Mol Phylogenet Evol* 109:180–190. <https://doi.org/10.1016/j.ympev.2017.01.004>
24. Nolasco-Soto J, Favila ME, Espinosa De Los Monteros A, et al (2020) Variations in genetic structure and male genitalia suggest recent lineage diversification in the Neotropical dung beetle complex *Canthon cyanellus* (Scarabaeidae: Scarabaeinae). *Biol J Linn Soc* 131:505–520. <https://doi.org/10.1093/biolinnean/blaa131>
25. Ratnasingham S, Hebert PDN (2013) A DNA-Based Registry for All Animal Species: The Barcode Index Number (BIN) System. *PLOS ONE* 8:e66213. <https://doi.org/10.1371/journal.pone.0066213>
26. Vieira MLC, Santini L, Diniz AL, et al (2016) Microsatellite markers: what they mean and why they are so useful. *Genet Mol Biol* 39:312–328. <https://doi.org/10.1590/1678-4685-GMB-2016-0027>
27. Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom
28. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17:10–12. <https://doi.org/10.14806/ej.17.1.200>
29. Bankevich A, Nurk S, Antipov D, et al (2012) SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>
30. Megléc E, Pech N, Gilles A, et al (2014) QDD version 3.1: a user-friendly computer program for microsatellite selection and primer design revisited: experimental validation of variables determining genotyping success rate. *Mol Ecol Resour* 14:1302–1313. <https://doi.org/10.1111/1755-0998.12271>
31. Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. *Nat Biotechnol* 18:233–234. <https://doi.org/10.1038/72708>
32. Guichoux E, Lagache L, Wagner S, et al (2011) Current trends in microsatellite genotyping. *Mol Ecol Resour* 11:591–611. <https://doi.org/10.1111/j.1755-0998.2011.03014.x>
33. Peakall R, Smouse PE (2006) genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes* 6:288–295. <https://doi.org/10.1111/j.1471-8286.2005.01155.x>
34. Liu K, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21:2128–2129. <https://doi.org/10.1093/bioinformatics/bti282>

35. Valière N (2002) gimlet: a computer program for analysing genetic individual identification data. *Mol Ecol Notes* 2:377–379. <https://doi.org/10.1046/j.1471-8286.2002.00228.x>
36. Amos W, Hoffman JI, Frodsham A, et al (2007) Automated binning of microsatellite alleles: problems and solutions. *Mol Ecol Notes* 7:10–14. <https://doi.org/10.1111/j.1471-8286.2006.01560.x>
37. Li CC (1976) *First Course in Population Genetics* Boxwood. Pac Grove CA
38. Manjeri G, Muhamad R, Faridah QZ, Tan SG (2014) Development of single locus DNA microsatellite markers in *Oryctes rhinoceros* (Linnaeus) using 5' anchored RAMs-PCR method. *J Genet* 93:92–96. <https://doi.org/10.1007/s12041-012-0189-8>
39. Song X, Yang T, Yan X, et al (2020) Comparison of microsatellite distribution patterns in twenty-nine beetle genomes. *Gene* 757:144919. <https://doi.org/10.1016/j.gene.2020.144919>
40. Behura SK, Severson DW (2015) Motif mismatches in microsatellites: insights from genome-wide investigation among 20 insect species. *DNA Res* 22:29–38. <https://doi.org/10.1093/dnares/dsu036>
41. Santana QC, Coetzee MPA, Steenkamp ET, et al (2009) Microsatellite discovery by deep sequencing of enriched genomic libraries. *BioTechniques* 46:217–223. <https://doi.org/10.2144/000113085>
42. Katti MV, Ranjekar PK, Gupta VS (2001) Differential Distribution of Simple Sequence Repeats in Eukaryotic Genome Sequences. *Mol Biol Evol* 18:1161–1167. <https://doi.org/10.1093/oxfordjournals.molbev.a003903>
43. Tóth G, Gáspári Z, Jurka J (2000) Microsatellites in Different Eukaryotic Genomes: Survey and Analysis. *Genome Res* 10:967–981. <https://doi.org/10.1101/gr.10.7.967>
44. Eder MLR, Rosa AL (2020) Non-tandem repeat polymorphisms at microsatellite loci in wine yeast species. *Mol Genet Genomics* 295:685–693. <https://doi.org/10.1007/s00438-020-01652-2>
45. Sampaio P, Correia A, Gusmão L, et al (2007) Sequence analysis reveals complex mutational processes for allele length variation at two polymorphic microsatellite loci in *Candida albicans*. In: *Communicating current research and educational topics and trends in applied microbiology*. Badajoz: Formatex. pp 926–935
46. Angers B, Bernatchez L (1997) Complex evolution of a salmonid microsatellite locus and its consequences in inferring allelic divergence from size information. *Mol Biol Evol* 14:230–238. <https://doi.org/10.1093/oxfordjournals.molbev.a025759>
47. Schlötterer C, Zangerl B (1999) The Use of Imperfect Microsatellites for DNA Fingerprinting and Population Genetics. In: Epplen JT, Lubjuhn T (eds) *DNA Profiling and DNA Fingerprinting*. Birkhäuser, Basel, pp 153–165
48. Viard F, Franck P, Dubois M-P, et al (1998) Variation of microsatellite size homoplasy across electromorphs, loci, and populations in three invertebrate species. *J Mol Evol* 47:42–51
49. Von Wahlund S (1928) Zusammensetzung von population und Korrelationsers cheimungen von Standpunkt der veierbungs lehreaus betachlet. *Hereditas* 11:65–106

Tables

Table 1. Mexican localities of *Canthon cyanellus* sampled in this study.

State/locality (code)	n	Coordinates (latitude, longitude)	Elevation (m a.s.l.)
Chiapas, El Vergel (Ver)*	2	14.7029, -92.2672	22
Chiapas, Palenque (Nbet)	2	17.2815, -91.6466	140
Guerrero, Ixtapa (Ixt)	2	17.6583, -101.5752	120
Jalisco, Chamela (Cha)	2	19.4997, -105.0229	90
Tamaulipas, Gómez-Farías (GF)	2	23.0480, -99.1433	379
Oaxaca, Huatulco (Hua)	2	15.7800, -96.0900	30
Veracruz, Tuxpan (Tp)*	3	20.9544, -97.4661	52
Veracruz, Los Tuxtlas (Tx)*	3	18.5833, -95.0667	120
* One sample from each of these localities was used for whole-genome sequencing for microsatellite loci discovery. n = sample size.			

Table 2. Contig assembly and microsatellite loci discovery in three samples of *Canthon cyanellus* genomic DNA.

Sample locality	El Vergel (Ver)	Los Tuxtlas (Tx)	Tuxpan (Tp)
Paired raw reads	6,872,992	8,252,350	6,059,490
Sequence length	35-251 bp	35-251 bp	35-251 bp
Paired reads after filtering	6,828,801 (99.4%)	8,207,156 (99.5%)	5,984,230 (98.8%)
Assembled contigs	89,037	115,242	113,636
N50	8,607 bp	6,037 bp	6,434 bp
N95	480 bp	441 bp	445 bp
Average contig length	2,922.3 bp	2,231.5 bp	2,263.8 bp
Minimum contig length	128 bp	128 bp	128 bp
Maximum contig length	87,324 bp	66,814 bp	76,419 bp
Total number of primer pairs	137,585	138,176	132,930
Total VNTR loci	9,794	9,705	9,333
Perfect microsatellites	8,718 (89.0%)	8,591 (88.5%)	8,315 (89.1%)
Final primer testing loci	0	8	8
VNTR = Variable number tandem repeat; Perfect microsatellites are defined as VNTR loci composed of a single, 2-6 bp long motif with no interruptions and at least five repetitions.			

Table 3. Characterization of 14 polymorphic microsatellite markers of *C. cyanellus*.

Locus	GenBank	Forward Primer	Reverse Primer	Motif	n	Size	A	A _E	H ₀	H _E
Ccy01	MW455319	GGGTATCTGTTGGGCACTAAT	TGATTGATTAATAACTGTACCGGCA	AAT	17	212-227	6	4.48	0.76	0.7
Ccy02	MW455320	CGATTGGGTTTGCTTTCAGAGA	AAAGCAACAACACTTGTCTACAT	AAT	18	96-108	5	2.10	0.39	0.5
Ccy03	MW455321	CTTCGTGGAAGAACAACCGA	ACTTAAACATTACAATTTGTTTGCCA	AAT	17	177-183	3	1.35	0.18	0.2
Ccy04	MW455322	AGTCGCGAAGAACTGACGAT	TCTTAATCGTAAATCGTGTTATGTACA	AAT	18	120-129	4	1.42	0.22	0.2
Ccy05	MW455323	ACAAATATGTGTCGTGCGCA	TCGAAATGTTCTCAGAATTCTTTG	AAT	17	304-310	4	2.42	0.53	0.5
Ccy06	MW455324	TCCTTTGGTGTTAACGACTTACA	TGAATCGATAGCGAAATAAGAGC	AAT	17	247-260	3	1.53	0.41	0.3
Ccy07	MW455325	TTGGCTAGTTGATCTGAAGGCA	GGCTGATTCTTGCCTTGACC	AAT	17	195-207	4	1.74	0.41	0.4
Ccy08	MW455326	TGGAATTTAATGATCAACCACGGA	TCGCCTCGATATCACCGAAA	AAT	16	145-151	4	2.40	0.19	0.5
Ccy09*	MW455327	ACTTTGTGATTTGTTATTGTGTACCT	TCTTGTGGTGGATTAGAACTGTCA	AAT	16	204-214	9	6.24	0.38	0.8
Ccy10	MW455328	TGATTGATTAATAACTGTACCGGCA	GGGTATCTGTTGGGCACTAAT	AAT	16	214-226	5	3.91	0.69	0.7
Ccy11	MW455329	GGCAACAATCTCAATCTTCTTATCTT	ACTAAAGTCTTTGATGATCTGCGT	AAT	18	188-197	3	1.25	0.11	0.2
Ccy12*	MW455330	ATCCAGCCAAACCGAATTCT	ACACACAAACAATAGAATTCTTCGA	AAT	17	243-267	16	12.04	0.82	0.9
Ccy13	MW455331	ACGCAAAGTCTATTGAAGGACA	ACTTTCTCCGAATGCCTTTACT	AAT	18	176-179	2	1.60	0.28	0.3
Ccy14	MW455332	AAGATCCCTTACGTAGTTGCGA	CGGACGCCAATAAACTCACC	AAG	18	142-151	4	2.54	0.33	0.6
Average						5.14		3.22	0.41	0.5
* = Loci that showed alleles with size differences shorter than three nucleotides; n = sample size; A = number of alleles; A _E = effective number of alleles; H ₀ = expected heterozygosity; F _{IS} = inbreeding coefficient; PIC = polymorphism information content; P _{ID} = probability of identity.										

Figures

Figure 1

Mexican localities of *Canthon cyanellus* sampled in this study. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Figure 2

Canthon cyanellus genomic characterization of variable number tandem repeat loci (VNTRs) using three samples. a) Proportions of three types of VNTR loci: P-MCs = Perfect Microsatellites; C-MCs = Compound Microsatellites; MNs = Minisatellites. b) Proportions of major VNTR loci classes; 2Nuc = Dinucleotides; 3Nuc = Trinucleotides; 4Nuc = Tetranucleotides; 5Nuc = Pentanucleotides; 6Nuc = Hexanucleotides. c) Number of reiteration units observed at each locus for various VNTRs loci classes in three *C. cyanellus* samples. Characterization of d) Dinucleotide and e) Trinucleotide VNTR loci by frequency of different motif types.

Figure 3

Binning the Ccy09 and Ccy12 alleles. Solid colours denote raw fragment lengths binned together into a single allele class.

Figure 4

Number of private alleles observed per locus per population.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryMaterial1.xlsx](#)
- [SupplementaryMaterial2.pdf](#)