

Genome Assembly Green Peach Aphid (GPA) (*Myzus persicae* Sulzer) clone O GPA_Ov1

The International Aphid Genomics Consortium (IAGC) is pleased to release the GPA clone O assembly version GPA_Ov1. Below you will find links to access the genome build and information relating to both its limitations and plans for improvement.

Pre-Publication Data-Access and Usage.

In order to promote public benefit arising from early access to the aphid genome assembly, this interim assembly (GPA_Ov1) is being made available prior to the publication of paper describing global analysis of the GPA genome. This prepublication data release is in accordance with the Bermuda and Fort Lauderdale agreements and the more recent Toronto Statement (Toronto International Data Release Workshop 2009), which provides guidelines for both data generators and data users. The IAGC asks that users adhere to each of these policies and refrain from publishing any global analysis of the GPA genome. Global analysis includes the identification of complete (whole genome) sets of genomic features such as genes, gene families, regulatory elements, repeat structures and GC content. Global analysis also includes chromosome wide or whole-genome scale comparisons with other species and reassemblies of the GPA data.

If you write a publication involving *Myzus persicae* sequences downloaded from this site, please provide the following statement: "DNA sequence data were downloaded from AphidBase, <http://www.aphidbase.com/aphidbase/>. Funding for *Myzus persicae* genomic sequencing was provided by The Genome Analyses Centre (TGAC) Capacity and Capability Challenge programme project CCC-15 and BB/J004553/1 from the Biotechnology and Biological Sciences Research Council (BBSRC) and the John Innes Foundation."

Technical specifications for GPA_Ov1.

GPA_Ov1 was built using next generation sequence derived from PGA clone O prevalent in the United Kingdom. The colony was started from a single parthenogenetically reproducing female. DNA extracted from aphids in this colony was used for paired-end (insert size: 500 bp), overlap paired-end (180 bp) and mate-pair (7.5 kb) sequencing. Primary de novo assembly was performed using Abyss 1.3.4 and scaffolds were generated with SSPACE 2.0 resulting in 34598 scaffolds with N50 of 148 kb, the longest scaffold of 1.02 Mb and total length of 398 Mb.

What is included in GPA_Ov1 and how to access the data.

This site has a BLAST functionality allowing users to search the assembly with input sequence of interest.

It is also possible to download single scaffolds.

This is considered a working draft release that contains both and unknown errors and discrepancies. We are working on improving this initial assembly and will provide this as an update on this site as soon as it becomes available.

What is scheduled for inclusion in PGA_Ov2.

- Provide gene model annotations of core set of conserved genes based on RNA-seq data
- Gene model predictions
- Comparative genome analyses between GPA clone O and G006.

The ISGC Reference Genome Project Work Team Is Led By:

Saskia Hogenhout (Department of Cell and Developmental Biology, John Innes Centre, Norwich, UK)

David Swarbreck (Genome Analysis Team, The Genome Analyses Centre, Norwich, UK)

Bernardo Clavijo (Scientific Programmer, The Genome Analyses Centre, Norwich, UK).