Figures and figure supplements

The genome of the Hi5 germ cell line from *Trichoplusia ni*, an agricultural pest and novel model for small RNA biology

*Yu Fu et al*
**Figure 1.** Chromosomes and genes in the *T. ni* genome based on data from the Hi5 cell line. (A) Genome assembly and annotation workflow. (B) An example of a DAPI-stained spread of Hi5 cell mitotic chromosomes used to determine the karyotype. (C) Phylogenetic tree and orthology assignment of *T. ni* with 18 arthropod and two mammalian genomes. Colors denote gene categories. The category 1:1:1 represents universal single-copy orthologs, allowing absence and/or duplication in one genome. N:N:N orthologs include orthologs with variable copy numbers across species, allowing absence in one genome or two genomes from different orders. Lepidoptera-specific genes are present in at least three of the four lepidopteran genomes; Hymenoptera-specific genes are present in at least one wasp or bee genome and at least one ant genome. Coleoptera-specific genes are present in both coleopteran genomes; Diptera-specific genes present in at least one fly genome and one mosquito genome. Insect indicates other insect-specific genes. Mammal-specific genes are present in both mammalian genomes. The phylogenetic tree is based on the alignment of 1:1:1 orthologs. DOI: https://doi.org/10.7554/eLife.31628.003
Figure 1—figure supplement 1. Hi5 cell Karyotyping. Thirty images showing the numbers of chromosomes (N) in Hi5 cells. N ranged from 103 to 122; mean ± S.D. = 111.7 ± 5.45. Since lepidopteran cell lines are typically tetraploid, the haploid genome likely contains 28 (mean ± S.D. = 27.9 ± 1.36) pairs of chromosomes.

DOI: https://doi.org/10.7554/eLife.31628.004
Figure 1—figure supplement 2. Phylogenetic tree of 21 species showing the scale, branch lengths and bootstrap support. Strict 1:1:1 orthologs were used to compute the phylogenetic tree using the maximum likelihood method. Black, branch length; red, bootstrap support.

DOI: https://doi.org/10.7554/eLife.31628.005
Figure 1—figure supplement 3. Opsins in insects.
DOI: https://doi.org/10.7554/eLife.31628.006
Figure 2. T. ni males are ZZ and females are ZW. (A) Normalized contig coverage in males and females. (B) Relative repeat content, gene density, transcript abundance (female and male thoraces), and piRNA density of autosomal, Z-linked, and W-linked contigs (ovary). (C) Multiple sequence alignment of the conserved region of the sex-determining gene *masc* among the lepidopteran species.

DOI: https://doi.org/10.7554/eLife.31628.008
Figure 2—figure supplement 1. T. ni sex determination and dosage compensation. (A) Genomic coverage comparison of Z-linked, W-linked and autosomal contigs. Contig coverage was shuffled 1,000,000 times to calculate the coverage ratio. Outliers are not shown. (B) Autosomal, Z-linked and W-linked transcript abundance in Hi5 cells and T. ni tissues. (C) Transcript abundance ratios of autosomal, Z-linked, and W-linked genes in Hi5 cells and T. ni tissues. (D) 95% confidence intervals for T. ni doublesex.
Figure 2—figure supplement 1 continued

*T. ni* tissues. Error bars represent 95% confidence interval estimated from 1000 bootstrap replicates. (D). Sex-specific splicing of *T. ni* doublesex pre-mRNA.

DOI: https://doi.org/10.7554/eLife.31628.009
Figure 2—figure supplement 2. CpG ratios and transposons. (A) Distribution of observed-to-expected CpG ratios in protein-coding genes (left panel) and in 500 bp genomic windows (right panel) in *A. mellifera*, *B. mori*, *D. plexippus*, *D. melanogaster*, *P. xylostella*, *T. castaneum*, and *T. ni*. (B) Proportion of the genome occupied by transposons versus transposon sequence divergence. Sequence divergence was calculated by comparing individual transposon copies with the corresponding consensus sequence (see Materials and methods). (C) Repeat content in lepidopteran genomes. DOI: https://doi.org/10.7554/eLife.31628.010
Figure 3. miRNA expression in T. ni. (A) Comparison of miRNA abundance in male and female T. ni thoraces. Solid circles, miRNAs with FDR < 0.1 and fold change >2. Outlined circles, all other miRNAs. (B) Comparison of the tissue distribution of the 44 most abundant miRNAs among T. ni ovaries, testes, and Hi5. (C) Heat map showing the abundance of miRNAs in (B). miRNAs are ordered according to abundance in ovary. Conservation status uses the same color scheme in (A).

DOI: https://doi.org/10.7554/eLife.31628.011
Figure 4. siRNA. (A) Distribution of siRNAs mapping to TNCL virus in the genomic (blue) and anti-genomic orientation (red). Inset: length distribution of TNCL virus-mapping small RNAs. (B) Distance between the 3’ and 5’ ends of siRNAs on opposite viral strands. (C) Distance between the 3’ and 5’ ends of siRNAs on the same strand. (D) Frequency of small RNA length in different tissues. (E) Log ratio of siRNA oxidized/unoxidized in D. melanogaster, P. xylostella, B. mori, T. ni, Diptera, and Lepidoptera.
ends of siRNAs on the same viral strand. (D) Length distribution of small RNAs from unoxidized and oxidized small RNA-seq libraries. (E) Lepidopteran siRNAs are not 2'-O-methylated. The box plots display the ratio of abundance (as a fraction of all small RNAs sequenced) for each siRNA in oxidized versus unoxidized small RNA-seq libraries. The tree shows the phylogenetic relationships of the analyzed insects. Outliers are not shown.

DOI: https://doi.org/10.7554/eLife.31628.012
Figure 4—figure supplement 1. T. ni siRNAs. (A) siRNA length distributions for multiple insects in oxidized and unoxidized small RNA-seq libraries. (B) Length distribution of fully matched and tailed TNCL virus-siRNAs.

DOI: https://doi.org/10.7554/eLife.31628.013

Fu et al. eLife 2018;7:e31628. DOI: https://doi.org/10.7554/eLife.31628
Figure 4—figure supplement 2. Loading asymmetry of siRNAs mapping to TNCL RNA1 (A) and RNA2 (B). For each single-stranded siRNA species, we searched for siRNAs on the other strand that when paired produce a typical siRNA duplex with two-nucleotide overhanging 3' ends. DOI: https://doi.org/10.7554/eLife.31628.014
Figure 5. piRNAs and miRNAs in the *T. ni* genome. (A) Abundance of mRNAs encoding piRNA pathway proteins in Hi5 cells, ovary, testis, and thorax. (B) Ideogram displaying the positions of miRNA genes (arrowheads) and piRNA clusters in the *T. ni* genome. Color-coding reports tissue expression for Figure 5 continued on next page
Hi5 cells, ovaries, testis, and thorax. Contigs that cannot be placed onto chromosome-length scaffolds are arbitrarily concatenated and are marked ‘Un.’ (C) Distribution of piRNAs among the autosomes, Z, and W chromosomes in Hi5 cells, ovary, testis, and female and male thorax, compared with the fraction of the genome corresponding to autosomes, W, and Z chromosomes.

DOI: https://doi.org/10.7554/eLife.31628.015
Figure 5—figure supplement 1. piRNA abundance (ppm) along the most productive piRNA cluster. Top, fixed scale (some data clipped); bottom, auto-scaled.

DOI: https://doi.org/10.7554/eLife.31628.016
Figure 5—figure supplement 2. T. ni piRNAs. (A) piRNA clusters tend to produce piRNAs that are antisense to transposons. The x-axis represents the ratio of piRNAs from the plus strand to piRNAs from the minus strand, with the dotted lines indicating twofold difference. The y-axis indicates the ratio of antisense transposon insertion.

(B) piRNA density across different tissues. Significance levels are indicated by asterisks. Hi5, Ovary, Testis, Female thorax, and Male thorax.

(C) piRNA abundance (ppm) across different categories of homology. Transposon homology (n = 8), Uncharacterized or hypothetical (n = 21), Potential protein-coding (n = 12), and No homologs (n = 33).
Figure 5—figure supplement 2 continued

of transposons lengths on the plus strand over transposon length on the minus strand. The solid line indicates regression line and shading indicates 95% confidence interval by LOWESS. Boxplot shows fractions of antisense transposons (i.e. transposons inserted opposite to the direction of piRNAs precursor transcription) in dual- and uni-strand clusters. Outliers are not shown. Wilcoxon rank-sum test. (B) piRNA densities on autosomal, Z-linked and W-linked contigs in Hi5 cells, ovary, testis, and female and male thorax. (C) Abundance of piRNAs from putative W-linked genes.

DOI: https://doi.org/10.7554/eLife.31628.017
Figure 6. *T. ni* piRNAs. (A) Hi5-specific piRNA clusters contain younger transposon copies. RC, rolling-circle transposons; LINE, Long interspersed nuclear elements; LTR, long terminal repeat retrotransposon; DNA, DNA transposon. (B) Comparison of piRNA abundance per cluster in female and male thorax. (C) piRNA precursors are rarely spliced. The number of introns supported by exon-exon junction-mapping reads is shown for protein-coding genes and for piRNA clusters for each tissue or cell type. (D) piRNA precursors are inefficiently spliced. Splicing efficiency is defined as the ratio of spliced over unspliced reads. Splice sites were categorized into those inside and outside piRNA clusters. Outliers are not shown.

DOI: https://doi.org/10.7554/eLife.31628.018
Figure 6—figure supplement 1. *T. ni* piRNA clusters. (A) Comparison of piRNA abundance (ppm) from ovary and Hi5 piRNA-producing loci and from ovary and testis piRNA-producing loci. (B) piRNA cluster lengths in *T. ni* ovary, testis, thorax, and Hi5 cells. (C) Motifs around intron boundaries of predicted protein-coding gene models within and outside of piRNA clusters.

DOI: https://doi.org/10.7554/eLife.31628.019
Figure 7. Genome editing in Hi5 cells. (A) Strategy for using Cas9/sgRNA RNPs to generate a loss-of-function TnPiwi deletion allele. Red, protospacer-adjacent motif (PAM); blue, protospacer sequence. Arrows indicate the diagnostic forward and reverse primers used in PCR to detect genomic deletions. (B) An example of PCR analysis of a TnPiwi deletion event. (C) Strategy for using Cas9/sgRNA RNPs and a single-stranded DNA homology donor to insert EGFP and an HA-tag in-frame with the vasa open-reading frame. (D) An example of PCR analysis of a successful HDR event. DNA isolated from wild type (WT) and FACS-sorted, EGFP-expressing Hi5 cells (HDR) were used as templates.

DOI: https://doi.org/10.7554/eLife.31628.020
Figure 8. Hi5 cells contain nuage. (A) Schematic of single-clone selection of genome-edited Hi5 cells using the strategy described in Figure 7C. (B) A representative field of Hi5 cells edited to express EGFP-HA-Vasa from the endogenous locus. (C) A representative image of a fixed, EGFP-HA-Vasa-expressing Hi5 cell stained with DAPI, anti-EGFP and anti-HA antibodies. EGFP and HA staining colocalize in a perinuclear structure consistent with Vasa localizing to nuage.

DOI: https://doi.org/10.7554/eLife.31628.021