**Supporting information**

**Table S1.** Metadata for all isolates included in this study: *X. fastidiosa* subspecies, *A. tumefaciens*, *X. oryzae*, *X. citri*, and *X. campestris*.

**Table S2.** Mixed effect ANOVA results on gene-specific GC content using original Roary’s accessory/core gene classification; original Roary’s accessory/core gene classification after removal of Hypothetical/Poorly characterized proteins; Roary's accessory/core gene classification using 95% as core genome cutoff; and GET\_HOMOLOGUES’s accessory/core gene classification.

**Table S3.** Mixed effect ANOVA results on gene-specific GC3 content.

**Table S4.** Mixed effect ANOVA results on gene-specific GC content after removal of ribosomal protein coding genes from the dataset.

**Table S5.** Mixed effect ANOVA results on gene-specific GC content using detailed gene functions groups.

**Table S6.** Core genome alignment length, number of SNPs, and Genetic diversity (π) of *X. fastidiosa*, *A. tumefaciens*, *X. oryzae*, *X. campestris*, and *X. citri*.

**Table S7.** List of genes with dN/dS>1 in subsp. *multiplex*, subsp. *pauca*, and subsp. *fastidiosa*. The corresponding dN/dS, genic GC content, r/m, and COG functional class are reported for each gene. Hypothetical genes are named after the group number assigned by Roary.

**Figure S1.** Boxplot showing marginal-GC in genes of the accessory genome organized by the number of orthologs. Each box represents different size of the accessory genome. Isolates have been divided by *X. fastidiosa* subspecies.

**Figure S2.** Boxplot showing marginal-GC in genes coded on the leading (light blue) and lagging (light red) DNA strand. Isolates have been divided by *X. fastidiosa* subspecies and by genes have been split into core/accessory genome components.

**Figure S3.** Boxplot showing GT in genes coded on the leading (light blue) and lagging (light red) DNA strand. Isolates have been divided by *X. fastidiosa* subspecies and by genes have been split into core/accessory genome components.

**Figure S4.** Boxplot showing marginal-GC content distribution on different functional classes after removal of ribosomal protein coding genes from datasets. Genes have been divided according to *X. fastidiosa* subspecies and core/accessory genome. Functional classes include: Information Storage and Processing (ISP, green), Cellular Signaling and Processing (CPS, blue), Metabolism (M, red), and Multiple categories (MU, purple).

**Figure S5.** Boxplot showing marginal-GC in detailed functional groups. Genes on the core genome are shown in light blue and genes in the accessory genome are shown in light red.

**Figure S6.** Boxplot showing gene size on different functional classes after removal of ribosomal protein coding genes. Genes have been divided according to *X. fastidiosa* subspecies and core/accessory genome. Functional classes include: Information Storage and Processing (ISP, green), Cellular Signaling and Processing (CPS, blue), Metabolism (M, red), and Multiple categories (MU, purple).

**Figure S7.** Synteny analysis for three finished *X. fastidiosa* chromosomes. Each plot shows comparisons between *X. fastidiosa* subsp. *fastidiosa* (strain Temecula 1), *X. fastidiosa* subsp. *multiplex* (strain M12), *X. fastidiosa* subsp. *pauca* (strain 9a5c). Several inversion events are observed between the 9a5c strain and the M12 and Temecula 1 strains. Analysis were performed using the web platform CoGe (<https://genomevolution.org/coge/>) and can be replicated following the listed links.

**Figure S8.** Boxplot showing marginal-GC content in recombinant vs. non-recombinant genes in three *X. fastidiosa* subspecies.

**Figure S9.** Boxplot showing gene size on different functional classes. Genes have been divided according to *X. fastidiosa* subspecies and recombinant/non-recombinant genes in the core genome. Functional classes include: Information Storage and Processing (ISP, green), Cellular Signaling and Processing (CPS, blue), Metabolism (M, red), and Multiple categories (MU, purple).

**Table S1.** Metadata for all isolates included in this study: *X. fastidiosa* subspecies, *A. tumefaciens*, *X. oryzae*, *X. citri*, and *X. campestris*.

See attached excel file.

**Table S2.** Mixed effect ANOVA results on gene-specific GC content using original Roary’s accessory/core gene classification; original Roary’s accessory/core gene classification after removal of Hypothetical/Poorly characterized proteins; Roary's accessory/core gene classification using 95% as core genome cutoff; and GET\_HOMOLOGUES’s accessory/core gene classification.

See attached excel file.

**Table S3.** Mixed effect ANOVA results on gene-specific GC3 content.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Subspecies** |  | **GC3** | | | | | | | | |
| **Genome** | | | | **Accessory** | | | **Core** | | |
| **Group** | **Chisq** | **Df** | **Pval** | **Chisq** | **Df** | **Pval** | **Chisq** | **Df** | **Pval** |
| *X. fastidiosa* subsp. *fastidiosa* | Accessory/Core | 1967.64 | 1 | **<2x10-16\*** | - | - | - | - | - | - |
| Host | 11.31 | 7 | **1.26x10-01\*** | 34.47 | 7 | **1.41x10-05\*** | 0.44 | 7 | 0.999 |
| Gene size | 13897.60 | 1 | **<2x10-16\*** | 5048.20 | 1 | **<2x10-16\*** | 6477.27 | 1 | **<2x10-16**\* |
| Strand | 430.37 | 1 | **<2x10-16\*** | 585.56 | 1 | **<2x10-16\*** | 3.69 | 1 | **5.47x10-02\*** |
| Population**†** | 4.28 | 4 | 0.369 | 9.14 | 4 | **5.78x10-02\*** | 0.10 | 4 | 0.999 |
| Function | 10339.52 | 4 | **<2x10-16**\* | 1967.65 | 4 | **<2x10-16\*** | 17903.19 | 4 | **<2x10-16**\* |
| Number of orthologs**‡** | 50140.99 | 182 | **<2x10-16\*** | 35980.03 | 182 | **<2x10-16\*** | - | - | - |
| *X. fastidiosa* subsp. *multiplex* | Accessory/Core | 710.39 | 1 | **<2x10-16**\* | - | - | - | - | - | - |
| Host | 22.53 | 18 | 0.209 | 25.80 | 18 | 0.104 | 1.19 | 18 | 1 |
| Gene size | 1286.79 | 1 | **<2x10-16**\* | 591.02 | 1 | **<2x10-16**\* | 875.45 | 1 | **<2x10-16**\* |
| Strand | 33.47 | 1 | **7.24x10-09\*** | 10.02 | 1 | **1.55x10-03\*** | 46.26 | 1 | **1.04x10-11\*** |
| Population**†** | 10.04 | 4 | 0.04 | 12.86 | 4 | 0.012 | 0.25 | 4 | 0.993 |
| Function | 1346.43 | 4 | **<2x10-16**\* | 213.47 | 4 | **<2x10-16\*** | 2315.16 | 4 | **<2x10-16**\* |
| Number of orthologs**‡** | 1468.82 | 38 | **<2x10-16\*** | 1165.34 | 38 | **<2x10-16\*** | - | - | - |
| *X. fastidiosa* subsp. *pauca* | Accessory/Core | 211.04 | 1 | **<2x10-16**\* | - | - | - | - | - | - |
| Host | 10.43 | 5 | 0.064 | 10.73 | 5 | 0.057 | 0.07 | 5 | 0.999 |
| Gene size | 8101.26 | 1 | **<2x10-16**\* | 6764.33 | 1 | **<2x10-16**\* | 1202.16 | 1 | **<2x10-16**\* |
| Strand | 33.32 | 1 | **7.83x10-06\*** | 22.59 | 1 | **2.00x10-06\*** | 11.50 | 1 | **6.95x10-04\*** |
| Population**†** | 43.30 | 4 | **8.98x10-09\*** | 43.97 | 4 | **6.51x10-09\*** | 6.97 | 4 | 0.137 |
| Function | 4972.89 | 4 | **<2x10-16\*** | 3740.81 | 4 | **<2x10-16\*** | 1601.44 | 4 | **<2x10-16**\* |
| Number of orthologs**‡** | 26053.32 | 103 | **<2x10-16\*** | 23457.20 | 103 | **<2x10-16\*** | - | - | - |

\***Statistically significant differences.**

**† Refers to the geographic populations included in this study: California, Southeastern USA, Taiwan, Spain, Brazil, Italy, Costa Rica, and France.**

**‡ Refers to the number of orthologues for each gene.**

**Table S4.** Mixed effect ANOVA results on gene-specific GC content after removal of ribosomal protein coding genes from the dataset.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Subspecies** | **GC** | | | | | | | | | |
| **Genome** | | | | **Accessory** | | | **Core** | | |
| ***X. fastidiosa* subsp. *fastidiosa*** | **Variable** | **Chisq** | **Df** | **Pr(>Chisq)** | **Chisq** | **Df** | **Pr(>Chisq)** | **Chisq** | **Df** | **Pr(>Chisq)** |
| Accessory/Core | 162.7784 | 1 | **<2x10-16\*** | - | - | **-** | - | - | **-** |
| Host | 155.458 | 8 | **<2x10-16\*** | 57.045 | 8 | **1.77x10-09\*** | 444 | 8 | **<2x10-16\*** |
| Gene length | 11549.092 | 1 | **<2x10-16\*** | 7456.2753 | 1 | **<2x10-16\*** | 383.48 | 1 | **<2x10-16\*** |
| Strand | 875.0346 | 1 | **<2x10-16\*** | 416.2369 | 1 | **<2x10-16\*** | 805.46 | 1 | **<2x10-16\*** |
| Population† | 9.9093 | 7 | 0.1938 | 4.6905 | 7 | 0.6977 | 20.01 | 7 | 0.0055 |
| Function | 1728.1135 | 4 | **<2x10-16\*** | 618.7177 | 4 | **<2x10-16\*** | 10099.9 | 4 | **<2x10-16\*** |
| Number of orthologs‡ | 6746.7097 | 189 | **<2x10-16\*** | 2458.901 | 189 | **<2x10-16\*** | - | - | **-** |
| ***X. fastidiosa* subsp. *multiplex*** | Accessory/Core | 16.5517 | 1 | **4.73x10-05\*** | - | - | **-** | - | - | **-** |
| Host | 13.544 | 15 | 0.5604 | 12.4927 | 15 | 0.6414 | 14.0804 | 15 | 0.5194 |
| Gene length | 419.0989 | 1 | **<2x10-16\*** | 255.1272 | 1 | **<2x10-16\*** | 97.9301 | 1 | **<2x10-16\*** |
| Strand | 56.6585 | 1 | **5.19x10-14\*** | 20.179 | 1 | **7.05x10-06\*** | 134.2912 | 1 | **<2x10-16\*** |
| Population† | 5.4511 | 5 | 0.3633 | 6.0869 | 5 | 0.2979 | 1.3441 | 5 | 0.9303 |
| Function | 272.0815 | 4 | **<2x10-16\*** | 196.8562 | 4 | **<2x10-16\*** | 81.2937 | 4 | **<2x10-16\*** |
| Number of orthologs‡ | 576.4062 | 37 | **<2x10-16\*** | 354.634 | 37 | **<2x10-16\*** | - | - | **-** |
| ***X. fastidiosa* subsp. *pauca*** | Accessory/Core | 55.7319 | 1 | **8.31x10-14\*** | - | - | **-** | - | - | **-** |
| Host | 2.4882 | 5 | 0.7783 | 0.7541 | 5 | 0.9799 | 1.0277 | 5 | 0.9603 |
| Gene length | 3593.3198 | 1 | **<2x10-16\*** | 3060.2776 | 1 | **<2x10-16\*** | 758.9033 | 1 | **<2x10-16\*** |
| Strand | 525.653 | 1 | **<2x10-16\*** | 457.0479 | 1 | **<2x10-16\*** | 76.3242 | 1 | **<2x10-16\*** |
| Population† | 1.4042 | 3 | 0.7046 | 1.2317 | 3 | 0.7454 | 3.3509 | 3 | 0.3406 |
| Function | 272.9506 | 4 | **<2x10-16\*** | 326.8554 | 4 | **<2x10-16\*** | 118.7646 | 4 | **<2x10-16\*** |
| Number of orthologs‡ | 3049.7275 | 103 | **<2x10-16\*** | 1954.636 | 103 | **<2x10-16\*** | - | - | **-** |

\***Statistically significant differences.**

**† Refers to the geographic populations included in this study: California, Southeastern USA, Taiwan, Spain, Brazil, Italy, Costa Rica, and France.**

**‡ Refers to the number of orthologues for each gene.**

**Table S5.** Mixed effect ANOVA results on gene-specific GC content using detailed gene functions groups.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Subspecies** | **GC** | | | | | | | | | |
| **Genome** | | | | **Accessory** | | | **Core** | | |
| ***X. fastidiosa* subsp. *fastidiosa*** | **Variable** | **Chisq** | **Df** | **Pr(>Chisq)** | **Chisq** | **Df** | **Pr(>Chisq)** | **Chisq** | **Df** | **Pr(>Chisq)** |
| Accessory/Core | 25.0773 | 1 | **5.51x10-07\*** | - | - | **-** | - | - | **-** |
| Host | 756.0221 | 8 | **<2x10-16\*** | 193.2607 | 8 | **<2x10-16\*** | 838.796 | 8 | **<2x10-16\*** |
| Gene length | 649.2673 | 1 | **<2x10-16\*** | 1444.0831 | 1 | **<2x10-16\*** | 471.013 | 1 | **<2x10-16\*** |
| Strand | 639.0637 | 1 | **<2x10-16\*** | 47.4737 | 1 | **5.58x10-12\*** | 665.108 | 1 | **<2x10-16\*** |
| Population† | 1.6802 | 7 | 0.9754 | 5.8146 | 7 | 0.5616 | 2.566 | 7 | 0.922 |
| Detailed function | 132.2292 | 51 | **3.89x10-09\*** | 54.6207 | 40 | 0.0615 | 164.222 | 41 | **<2x10-16\*** |
| Number of orthologs‡ | 6.814 | 1 | **0.0090\*** | 10.2443 | 1 | **0.0014\*** | - | - | **-** |
| ***X. fastidiosa* subsp. *multiplex*** | Accessory/Core | 29.584 | 1 | **5.35x10-08\*** | - | - | **-** | - | - | **-** |
| Host | 15.006 | 15 | 0.451 | 17.3172 | 15 | 0.3003 | 67.2117 | 15 | **1.40x10-08\*** |
| Gene length | 392.594 | 1 | **<2x10-16\*** | 252.0168 | 1 | **<2x10-16\*** | 6.1093 | 1 | **0.0135\*** |
| Strand | 97.969 | 1 | **<2x10-16\*** | 43.1517 | 1 | **5.07x10-11\*** | 72.7471 | 1 | **<2x10-16\*** |
| Population† | 1.007 | 5 | 0.962 | 4.8475 | 5 | 0.4348 | 7.4748 | 5 | 0.1877 |
| Detailed function | 3501.883 | 49 | **<2x10-16\*** | 2163.865 | 41 | **<2x10-16\*** | 160.5433 | 35 | **<2x10-16\*** |
| Number of orthologs‡ | 146.732 | 1 | **<2x10-16\*** | 109.5167 | 1 | **<2x10-16\*** | - | - | **-** |
| ***X. fastidiosa* subsp. *pauca*** | Accessory/Core | 0.4637 | 1 | 0.4959 | - | - | **-** | - | - | **-** |
| Host | 0.3554 | 5 | 0.9965 | 11.156 | 5 | **0.0484\*** | 3.8539 | 5 | 0.5706 |
| Gene length | 1262.752 | 1 | **<2x10-16\*** | 150.228 | 1 | **<2x10-16\*** | 1799.1638 | 1 | **<2x10-16\*** |
| Strand | 556.8399 | 1 | **<2x10-16\*** | 416.995 | 1 | **<2x10-16\*** | 134.1542 | 1 | **<2x10-16\*** |
| Population† | 1.881 | 3 | 0.5975 | 21.275 | 3 | **9.23x10-05\*** | 2.3537 | 3 | 0.5023 |
| Detailed function | 202.4666 | 51 | **<2x10-16\*** | 412.471 | 46 | **<2x10-16\*** | 57.1757 | 31 | **0.0029\*** |
| Number of orthologs‡ | 27.9296 | 1 | **1.26x10-07\*** | 19.581 | 1 | **9.64x10-06\*** | - | - | **-** |

\***Statistically significant differences.**

**† Refers to the geographic populations included in this study: California, Southeastern USA, Taiwan, Spain, Brazil, Italy, Costa Rica, and France.**

**‡ Refers to the number of orthologues for each gene.**

**Table S6.** Core genome alignment length, number of SNPs, and Genetic diversity (π) of *X. fastidiosa*, *A. tumefaciens*, *X. oryzae* pv*. oryzae, X. oryzae* pv*. oryzicola*, *X. campestris*, and *X. citri*.

|  |  |  |  |
| --- | --- | --- | --- |
| **Population** | **SNPs** | **Core (nt)** | **Nucleotide diversity (π)** |
| *X. fastidiosa* subsp. *fastidiosa* (N=194) | 15,646 | 1,378,202 | 0.0008 |
| *X. fastidiosa* subsp. *multiplex* (N=39) | 5,347 | 770,877 | 0.0011 |
| *X. fastidiosa* subsp. *pauca* (N=107) | 2,393 | 1,185,478 | 0.0004 |
| *X. oryzae pv. oryzae* (N=75) | 45,414 | 2,541,773 | 0.007 |
| *X. oryzae* pv*. oryzicola* (N=13) | 22,029 | 3,351,890 | 0.002 |
| *X. citri* (N=68) | 111,198 | 1,138,072 | 0.012 |
| *X. campestris* (N=14) | 113,771 | 3,485,626 | 0.011 |
| *A. tumefaciens* (N=12) | 283,876 | 1,699,729 | 0.059 |

**Table S7.** List of genes with dN/dS>1 in subsp. *multiplex*, subsp. *pauca*, and subsp. *fastidiosa*. The corresponding dN/dS, genic GC content, r/m, and COG functional class are reported for each gene. Hypothetical genes are named after the group number assigned by Roary.

See attached excel file.

**Supplementary figures**



**Figure S1.** Boxplot showing marginal-GC in genes of the accessory genome organized by the number of orthologs. Each box represents different size of the accessory genome. Isolates have been divided by *X. fastidiosa* subspecies.



**Figure S2.** Boxplot showing marginal-GC in genes coded on the leading (light blue) and lagging (light red) DNA strand. Isolates have been divided by *X. fastidiosa* subspecies and by genes have been split into core/accessory genome components.



**Figure S3.** Boxplot showing GT in genes coded on the leading (light blue) and lagging (light red) DNA strand. Isolates have been divided by *X. fastidiosa* subspecies and by genes have been split into core/accessory genome components.



**Figure S4.** Boxplot showing marginal-GC content distribution on different functional classes after removal of ribosomal protein coding genes from datasets. Genes have been divided according to *X. fastidiosa* subspecies and core/accessory genome. Functional classes include: Information Storage and Processing (ISP, green), Cellular Signaling and Processing (CPS, blue), Metabolism (M, red), and Multiple categories (MU, purple).



**Figure S5.** Boxplot showing marginal-GC in detailed functional groups. Genes on the core genome are shown in light blue and genes in the accessory genome are shown in light red.



**Figure S6.** Boxplot showing gene size on different functional classes after removal of ribosomal protein coding genes. Genes have been divided according to *X. fastidiosa* subspecies and core/accessory genome. Functional classes include: Information Storage and Processing (ISP, green), Cellular Signaling and Processing (CPS, blue), Metabolism (M, red), and Multiple categories (MU, purple).



**Figure S7.** Synteny analysis for three finished *X. fastidiosa* chromosomes. Each plot shows comparisons between *X. fastidiosa* subsp. *fastidiosa* (strain Temecula 1), *X. fastidiosa* subsp. *multiplex* (strain M12), *X. fastidiosa* subsp. *pauca* (strain 9a5c). Several inversion events are observed between the 9a5c strain and the M12 and Temecula 1 strains. Analysis were performed using the web platform CoGe (<https://genomevolution.org/coge/>) and can be replicated following the listed links.



**Figure S8.** Boxplot showing marginal-GC content in recombinant vs. non-recombinant genes in three *X. fastidiosa* subspecies



**Figure S9.** Boxplot showing gene size on different functional classes. Genes have been divided according to *X. fastidiosa* subspecies and recombinant/non-recombinant genes in the core genome. Functional classes include: Information Storage and Processing (ISP, green), Cellular Signaling and Processing (CPS, blue), Metabolism (M, red), and Multiple categories (MU, purple).