11.12.14

План на ближайшие пару дней:

nmpdr по-прежнему не работает, так что вопрос о том, откуда доставать инфу об эссенц+положении, все еще актуален)

Из DEG – там есть и эсс, и неэсс, и GI для сопоставления с генбанком.

краткое описание каждого нового эксперимента

Таблицы для всех, так или иначе

+ Скрипт для парсинга генбанка и определения цепи – найден

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Организм | Тип | Тип эксперимента | Условия выращивания | Статья |
| *Escherichia coli K-12 MG1655* | *Proteobacteria* | целенаправленная делеция | к | [40] |
| *Escherichia coli K-12*  *BW25113* | *Proteobacteria* | случайная вставка | п | [41] |
| *Helicobacter pylori G27* | *Proteobacteria* | случайная вставка | п | [42] |
| *Haemophilus influenzae KW20 Rd* | *Proteobacteria* | случайная вставка | п | [43] |
| *Salmonella enterica serovar. Typhimurium LT2* | *Proteobacteria* | случайная вставка | к | [44] |
| *Pseudomonas aeruginosa PA14* | *Proteobacteria* | случайная вставка | к | [45] |
| *Bacillus subtilis 168* | *Firmicutes* | целенаправленная вставка | к | [31] |
| *Streptococcus pneumoniae R6* | *Несколько ори!*  *Firmicutes* | целенаправленная вставка, целенаправленная делеция | к | [46], [47] |
|  |  |  | П? | [8] |
| *Mycoplasma genitalium* | *Firmicutes* | случайная вставка | к | [48] |
| *Staphylococcus aureus N315* | *Несколько ори!*  *Firmicutes* | случайный сайленсинг, целенаправленная делеция | к | [49]*,* [50] |
| *Mycobacterium tuberculosis H37Rv* | *Actinobacteria* | случайная вставка | п | [51] |
| Acinetobacter baylyi ADP1 | *В ncbi и ori – sp,but also ADP1* |  |  |  |
| Bacteroides thetaiotaomicron VPI-5482 |  |  |  |  |
| Burkholderia thailandensis E264 | *2 хромосомы* |  |  |  |
| Caulobacter crescentus |  |  |  |  |
| Francisella novicida U112 | A total of 312 predicted F. novicida genes lacked transposon insertions and therefore appear likely to be essential for growth on nutrient media. Other candidate essential genes are those with insertions only near the termini of the ORFs (the 3′ ends of some essential genes may be functionally dispensable, and some 5′ ends may tolerate insertion if transcription can restart or proceed through the transposon) and large genes with very few insertions (which could be due to insertions in tandemly duplicated genes or at rare permissive sites). -> 396 ess genes | Случайная вставка | клон | [7] |
| Mycoplasma pulmonis UAB CTIP | *Расширила ори и тер на 2000 (а не на 5000, как у остальных)*  *Около 150 тех, что есть в гб, в дег отсутствуют* |  |  |  |
| Porphyromonas gingivalis ATCC 33277 | putatively essential for viability in vitro on blood agar plates  highly-saturated mutant libraries  nested semi-random PCR followed by sequencing which confirmed that insertions occurred in multiple locations across the genome  Insertions into the **first or last five percent** of a gene have a higher likelihood of generating a functional gene product relative to insertions in the middle portions of a gene, therefore these mutants were **eliminated from consideration**.  a minimum of three sequencing ‘reads’ of the mutant locus be present in both technical replicates to exclude nonexistent insertions introduced by mapping of incorrectly sequenced reads, and lower rates may be due to mis-assignment by the reference assembly software   1,639 genes are non-essential for growth in vitro  The rules are similar to those used in previous essential gene analyses of other bacteria and contend that [[27](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3547785/#B27),[29](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3547785/" \l "B29),[30](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3547785/#B30),[42](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3547785/#B42)]: 1) A gene must contain at least 10 TA sites. Genes with less sites could be under-inserted due to random chance. Of the 204 genes in the P. gingivalis ATCC 33277 genome with less than 10 TA sites, 189 (93%) are annotated as hypothetical. 2) Genes found to have an actual to theoretical insertion ratio of 50-fold or greater under-insertion were considered putatively essential (actual:theoretical ≤ 0.020). Applying these rules, out of a total 2,102 genes in the ATCC 33277 genome (all protein coding sequences and rRNA genes combined minus the 53 tRNAs), we identified 463 (22.0%) genes as putatively essential for in vitro survival  **(совпадает с тем, что есть в deg;**  **1627 в dneg против 1639 здесь. rRNA в deg нет)**  **DEG записал в nonessential те гены, которые авторы исключили из рассмотрения как имеющие слишком мало потенциальных сайтов! Осторожнее с DNEG.**  **Данные об ess and noness, то и другое разумным образом проверено** | случайная вставка | клон | [4] |
| Salmonella enterica subsp. enterica serovar Typhimurium str. 14028S |  |  |  |  |
| Streptococcus sanguinis | *2 ori, в качестве ори задано пространство от начала первого до конца второго (1724 нукл)* |  |  |  |
| Vibrio cholerae N16961 | *2 хромосомы, в гбпримерно на 200 записей меньше, чем в деге оО* |  |  |  |
| Shewanella oneidensis *Нет в DEG* |  |  |  |  |
| transposon insertion frequency analysis (TIFA): For each gene, the expected number of insertions was estimated using a probability generating function with the sequence specific probabilities that were weighted by the genome locational specific bias. Using this probability model, the number of expected insertions was calculated for each gene and compared to the observed number of insertions ([Fig. 2](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4168976/figure/pcbi-1003848-g002/)). Genes were called essential if the combined probability of finding as few as, or fewer insertions than observed, was less than one over the number of genes in the dataset i.e. we accepted one false positive in our essential gene selection. For 50 of the **273 identified essential genes** a fitness value larger than zero was observed  The expected number of insertions for each gene was calculated using this essential gene model, and each gene that contained significantly more insertions than expected, was called nonessential. Following the earlier logic, a cutoff value was used that allowed for a single false-positive nonessential gene identification. This method identified **2,216 genes as nonessential**.  No essentiality call could be made for 1,722 genes, and three genes (SO2148, SO3175 and SO3872) were identified as both essential and nonessential (в зависимости от того, куда попала вставка)  Gene essentiality calls that were performed in previous work on the bases of absence or presence of insertions [[18]](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4168976/#pcbi.1003848-Deutschbauer1) were compared to TIFA calls in detail ([Table 1](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4168976/table/pcbi-1003848-t001/) and Dataset S1). Seventy eight essential gene calls and 1,958 nonessential gene calls were in agreement. Eighty one genes that were previously identified as nonessential (including two orthologs to essential genes in E. coli) were identified as essential by TIFA, 36 of which contained no insertions in our dataset. Fifty one of the genes previously identified as essential were identified as nonessential by TIFA, and contained an average of 10 insertions per gene  **Polar выкидываются из рассмотрения**  Кроме того, тут есть **данные по FBA** (flux balance analysis) essentiality – предсказание на основе метаболических карт  Плюс есть еще количественно измеренный фитнес (в основном для необязательных генов, тк нужно было найти хотя бы 8 вставок в обеих временных точках) – соотношение числа животных со вставкой до роста и после него | Случайная вставка | Популяция! | [5] |
|  |  | П, разн усл? | [6] |
| Bacteroides fragilis 638R | Massive parallel sequencing of **saturated transposon mutant librarie**s. mutant generation by the  transposon vector and identification of the transposon  disrupted genes is reproducible and reliable  550 genes (12.7%) were found to be essential for the survival  . Genes that had transposon insertions in  the **last 5%** of the gene (3' end) were filtered out since  they may likely to generate active product  Mariner transposons preferentially insert into  TA sites, therefore, we disregarded genes which have  **either less than 10 TA sites or were less than 150 bp** in  length, since these genes are likely to escape random  transposon disruption  Genes were c**onsidered essential if they were not**  **disrupted** by the transposon in either biological replicate.  Of the 3763 disrupted genes, 201 were  disrupted only once in either one or both the biological  replicates. Closer examination of these 201 mutant  reads indicated that transposon was integrated well  within the genes. In addition, all 201 genes were disrupted  in a mutant pool which was sequenced following  re-growth of mutants on BHI medium, confirming  that these genes are not essential for survival of B. fragilis  638R. – что подтверждает неоправданность идеи определять эссенциальность по количеству вставок  Есть данные по **экспрессии**  - см в статьях, приложение 1.  Насчет клонального выращивания: они смешивают культуру с еколи, (не отмечено никакого периода совместного выращивания), потом рассеивают на чашки, убивая еколи и фрагилисов без вставки. Похоже, растет не ковер, а отдельные колонии, поскольку в статье указано их примерное количество. | Случайная вставка | Клон | [Bfrag] |
| Burkholderia pseudomallei K96243 | **sufficient insertion density** for the application of robust statistics to identify genes that are essential for the in vitro growth  a library of over 1 million bacterial mutants was constructed using a modified miniTn5 transposon  TraDIS (12). Briefly, genomic DNA was isolated from two biological replicates representing separate cultures of the entire pool of mutants and sequenced using a primer specific to the 5= end of the transposon and reading directly into the surrounding genome sequence  .**Разница между хромосомами:** efrequency  of transposon insertion was notably higher in chromosome 1 than in chromosome 2  Approximately 170,000 of the insertion sites were mapped to chromosome 1, resulting in an average of 1 transposon insertion every 25 bp, while only approximately 70,000 mapped to chromosome 2, resulting in 1 insertion every 45 bp  we predicted that genes **with no or very few**  **insertion sites** are likely to be essential genes  we analyzed the number of insertion sites per gene after normalizing for gene length to create a gene insertion index. Performance of a  density estimate of the frequency distribution of gene insertion indexes results in a bimodal curve in which the first sharp peak represents genes in which a transposon insertion would be lethal to the bacteria and the second elongated peak represents genes which can be mutated without affecting the viability of the bacteriumThe gamma distributions of the density plot were used to estimate log2 likelihood ratios to distinguish essential from nonessential genes.  TraDIS screening showed only *one transposon insertion site within the sodB sequence*, compared to 68 unique insertion sites in the surrounding intergenic regions (Fig. 4D). Thus, our windowed algorithm determined that sodB is *predicted to be essential* in B. pseudomallei. Using the conditional lethal sodB mutant of B. pseudomallei, *we were able to confirm that the mutant was unable to grow* in the presence of glucose, and thus sodB is an essential gene – а вот тут для одного гена с одной вставкой подтверждена его эссенциальность  **Ничего про полярность, концы или мелкие гены (последнее понятно, они же делали поправку на размер)** |  |  |  |
| Campylobacter jejuni subsp. jejuni NCTC 11168 = ATCC 700819 | Есть 2 статьи, одна недоступна  metabolic model of C. jejuni was constructed using the annotation of the NCTC 11168 genome  sequence, a published model of the related bacterium Helicobacter pylori, and extensive literature mining. Using  this model, we have used **in silico Flux Balance Analysis** (FBA) to determine key metabolic routes that are essential  for generating energy and biomass, thus creating a list of genes potentially essential for growth under laboratory  conditions. To complement this in silico approach, **candidate essential genes have been determined using a whole genome transposon mutagenesis method**. F |  |  |  |
|  |  |  |  |  |
| Sphingomonas wittichii RW1 |  |  |  |  |

|  |
| --- |
| ORGANISM |
| Acinetobacter baylyi ADP1 |
| Bacillus subtilis 168 |
| Bacteroides thetaiotaomicron VPI-5482 |
| Burkholderia thailandensis E264 |
| Caulobacter crescentus |
| Escherichia coli MG1655 |
| Escherichia coli MG1655 I |
| Escherichia coli MG1655 II |
| Francisella novicida U112 |
| Haemophilus influenzae Rd KW20 |
| Helicobacter pylori 26695 |
| Mycobacterium tuberculosis H37Rv |
| Mycobacterium tuberculosis H37Rv II |
| Mycoplasma genitalium G37 |
| Mycoplasma pulmonis UAB CTIP |
| Porphyromonas gingivalis ATCC 33277 |
| Pseudomonas aeruginosa UCBPP-PA14 |
| Salmonella enterica serovar Typhi |
| Salmonella enterica subsp. enterica serovar Typhimurium str. 14028S |
| Salmonella typhimurium LT2 |
| Staphylococcus aureus N315 |
| Staphylococcus aureus NCTC 8325 |
| Streptococcus pneumoniae |
| Streptococcus sanguinis |
| Vibrio cholerae N16961  На самом деле их больше – см в деге   |  | | --- | | Acinetobacter baylyi ADP1 | | Bacillus subtilis 168 | | Bacteroides fragilis 638R | | Bacteroides thetaiotaomicron VPI-5482 | | Burkholderia pseudomallei K96243 | | Burkholderia thailandensis E264 | | Campylobacter jejuni subsp. jejuni NCTC 11168 = ATCC 700819 | | Caulobacter crescentus | | Escherichia coli MG1655 I | | Escherichia coli MG1655 II | | Francisella novicida U112 | | Haemophilus influenzae Rd KW20 | | Helicobacter pylori 26695 | | Mycobacterium tuberculosis H37Rv | | Mycobacterium tuberculosis H37Rv II | | Mycobacterium tuberculosis H37Rv III | | Mycoplasma genitalium G37 | | Mycoplasma pulmonis UAB CTIP | | Porphyromonas gingivalis ATCC 33277 | | Pseudomonas aeruginosa PAO1 | | Pseudomonas aeruginosa UCBPP-PA14 | | Salmonella enterica serovar Typhi | | Salmonella enterica serovar Typhi Ty2 | | Salmonella enterica serovar Typhimurium SL1344 | | Salmonella enterica subsp. enterica serovar Typhimurium str. 14028S | | Salmonella typhimurium LT2 | | Shewanella oneidensis MR-1 | | Sphingomonas wittichii RW1 | | Staphylococcus aureus N315 | | Staphylococcus aureus NCTC 8325 | | Streptococcus pneumoniae | | Streptococcus sanguinis | | Vibrio cholerae N16961 | |

Есть еще Эшерихии?

Файл BasicData

Лист Psae

Зеленым – данные SEED, deduced by comparison of the two gene essentiality datasets obtained independently in the P. aeruginosa strains PA14 [1] and PAO1 [2]

Голубым – данные OGEE, только [1]. Зато тут есть реальные имена локусов PA14, по которым можно найти соответствующую запись в генбанке. Впрочем, я верю тому соответствию, которое установили SEED

Персиковым: [3] Еще есть статья по PAO1 – полная коллекция мутантов, полученная вставкой транспозона. Тут содержатся информация обо всех генах, не являющихся необходимыми.

Все чушь про персиковое: это проверенные мутанты из статьи [2], так что нового тут только вот что: если ген назван неэссенциальным в этой статье, но можно быть увереннымЖ так оно и есть. Проверив пару генов, мутанты по которым есть в их библиотеке, и увидев, что они как раз проверены не самым лучшим образом, я с легким сердцем выкидываю данные этой статьи. Персиковое стало голубым)

В таблице BasicData Direction – genbank tag, Strand – 1 if the gene is on leading strand

Если инфа не используется в финальной таблице, она помечена голубым. Основная инфа – зеленым, дополнительная - персиковым

Исключаем из рассмотрения potential\_essential

Отступ от ори и тер – 5000 в каждую сторону.

[5] The presented transposon insertion frequency analysis (TIFA) improves on the more direct interpretation in which presence of an insertion in the gene core is interpreted as sufficient evidence of nonessentiality, and absence of insertions is interpreted as essentiality [[18]](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4168976/#pcbi.1003848-Deutschbauer1), [[25]](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4168976/#pcbi.1003848-Zomer1). The need for a more sophisticated approach has been recognized by others, with recent estimations of gene essentiality using hidden Markov models (HMM)

To investigate if mutants with insertions in genes that were called essential grew more slowly than other clones, the nonessentiality probability was plotted against fitness ([Fig. 3A](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4168976/figure/pcbi-1003848-g003/)). Although the average fitness of clones with insertions in essential genes was lower than fitness of clones with insertions in nonessential genes, the spread was very large, indicating that slow growth was an insufficient explanation for all detected insertions in essential genes.

Shon

Два исследования

Везде и всюду данные об эссенциальности приведены к виду – essential\nonessential\любая другая информация, которая означает – мы не знаем

Внимание!

Bacillus subtilis, Streptococcus (оба вида) и Staphylococcus – с несколькими ори! Для одного из них я вовремя это отловила, остальных еще не обрабатывала.

Программка на джаве не умеет правильно обращаться с такими хромосомами, но отлавливает их и выписывает в консоль.

Может быть, попробовать на сальмонеллах проверить, тяготеют ли гены под давлением движущего отбора (dn/ds) к запаздывающей? Нужно сопоставление всех генов

Почему именно на сальмонеллах? Нам не нужны для этого данные по необходимости генов. Хотя любопытно было бы посмотреть, есть ли необходимые гены под давлением движ отбора..

Дофига геномов туберкулеза вот у нас тут есть.

Если брать за основу гипотезу о повышенной мутабельности, нас интересуют не столько эссенциальные, сколько консервативные (именно по последовательности) гены. Есть ведь изменчивые эссенциальные гены, правда? Можно посмотреть на ряд ортологов, оценить его консервативность (нормировав на среднюю консервативность гена на соответствующей цепи), посмотреть, как часто в разных организмах ген попадает на лидирующую цепь.

Это некий аналог обратного предложения – посмотреть, куда идут гены с большим dn/ds

### [5] Mutant fitness and gene expression are poorly correlated

[1] [Liberati NT, Urbach JM, Miyata S, Lee DG, Drenkard E, Wu G, Villanueva J, Wei T, Ausubel FM: An ordered, nonredundant library of Pseudomonas aeruginosa strain PA14 transposon insertion mutants. Proc Natl Acad Sci U S A 2006.](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=16477005" \t "_blank)

[2] [Jacobs MA, Alwood A, Thaipisuttikul I, Spencer D, Haugen E, Ernst S, Will O, Kaul R, Raymond C, Levy R, et al.: Comprehensive transposon mutant library of Pseudomonas aeruginosa. Proc Natl Acad Sci U S A 2003, 100:14339-14344.](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=14617778)

[3] Held, Kiara et al. “Sequence-Verified Two-Allele Transposon Mutant Library for Pseudomonas Aeruginosa PAO1.” *Journal of Bacteriology* 194.23 (2012): 6387–6389. *PMC*. Web. 11 Dec. 2014.

[4] Klein BA, Tenorio EL, Lazinski DW, Camilli A, Duncan MJ, Hu LT. Identification of essential genes of the periodontal pathogen *Porphyromonas gingivalis*. *BMC Genomics* 2012;13:578. doi:10.1186/1471-2164-13-578.

[5] Yang H, Krumholz EW, Brutinel ED, et al. Genome-Scale Metabolic Network Validation of *Shewanella oneidensis* Using Transposon Insertion Frequency Analysis. Maranas CD, ed. *PLoS Computational Biology* 2014;10(9):e1003848. doi:10.1371/journal.pcbi.1003848.

[6]Deutschbauer A, Price MN, Wetmore KM, et al. Evidence-Based Annotation of Gene Function in *Shewanella oneidensis* MR-1 Using Genome-Wide Fitness Profiling across 121 Conditions. Richardson PM, ed. *PLoS Genetics*2011;7(11):e1002385. doi:10.1371/journal.pgen.1002385.

[7]Gallagher LA, Ramage E, Jacobs MA, Kaul R, Brittnacher M, Manoil C. A comprehensive transposon mutant library of *Francisella novicida*, a bioweapon surrogate. *Proceedings of the National Academy of Sciences of the United States of America* 2007;104(3):1009-1014. doi:10.1073/pnas.0606713104.

## [8] Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms

Tim van Opijnen[1](http://www.nature.com/nmeth/journal/v6/n10/full/nmeth.1377.html#a1), Kip L Bodi[1](http://www.nature.com/nmeth/journal/v6/n10/full/nmeth.1377.html#a1) & Andrew Camilli[1](http://www.nature.com/nmeth/journal/v6/n10/full/nmeth.1377.html#a1)

*Nature Methods* **6**, 767-772 (2009)   
Published online: 20 September 2009 | doi:10.1038/nmeth.1377

[Bfrag] Veeranagouda Y, Husain F, Tenorio EL, Wexler HM. Identification of genes required for the survival of *B. fragilis* using massive parallel sequencing of a saturated transposon mutant library. *BMC Genomics* 2014;15(1):429. doi:10.1186/1471-2164-15-429.

Identification of essential genes in *C. jejuni* genome highlights hyper-variable plasticity regions