

# Dirty Daphnia and Clean Ctenophore

Tatiana Lappi, Mila Zudina, Irena Artamonova



## Введение

- При современном секвенировании полных геномов возникает ряд проблем:
  - У секвенируемых организмов могут быть симбионты, чей генетический материал неминуемо попадает в пробу для секвенирования
  - При подготовке экспериментальной пробы возможны и другие источники загрязнения генетическим материалом других организмов (см. рис. 1)
  - Размеры секвенируемых участков ДНК ограничены по длине (чтения, ~ 70-1000 н.), в результате чего требуется дальнейшая сборка полученных последовательностей.
  - Чтения собираются в контиги, которые затем собираются в скаффолды. При сборке велика вероятность ошибочного расположения непрерывно прочитанных фрагментов относительно друг друга
  - В процессе секвенирования могут происходить ошибки и встречаются непрочитанные участки известной длины. Это связано с техническими особенностями процесса и наличием сложных для секвенирования участков ДНК
- Ранее был разработан метод выявления бактериального загрязнения в недособранных геномах эукариотических организмов. С помощью этого метода был описан облигатный эндосимбионт *Hydra magnipapillata*, доказано существование бактериальных эндосимбионтов *Nematostella vectensis*

## Цели

- Проверить, есть ли бактериальное загрязнение в недавно секвенированных геномах гребневика (*Pleurobrachia bachei*) и дафнии (*Daphnia pulex*) (см. рис. 2).

## Выводы

- В последовательности генома гребневика *Pleurobrachia bachei* мы не обнаружили бактериального загрязнения.
- В последовательности генома дафнии *Daphnia pulex* мы обнаружили бактериальное загрязнение несколькими видами бактерий, преимущественно - неописанным ранее видом из семейства Comamonadaceae, Burkholderiales, Betaproteobacteria и неописанным ранее представителем рода *Pseudomonas* (Gammaproteobacteria). Предположительно, обнаруженные бактерии являются симбионтами *Daphnia pulex*.

## Материалы и методы

- Методика исследования БПБП *Pleurobrachia bachei* и *Daphnia pulex* иллюстрирована блок-схемой на рис. 3
- Методика проверки наличия интронов в генах, кодирующих БПБП, проиллюстрирована на рис. 4

## Результаты

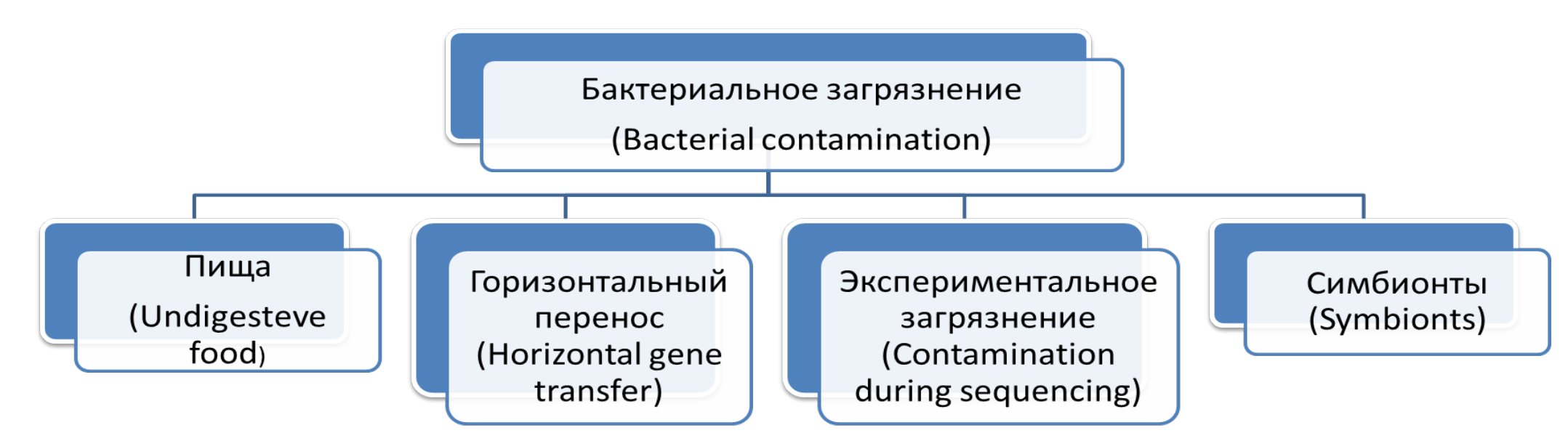


Рис. 1 Источники бактериального загрязнения в эукариотических геномных проектах  
Fig. 1 Bacterial contamination sources in the eukaryotic genomic projects

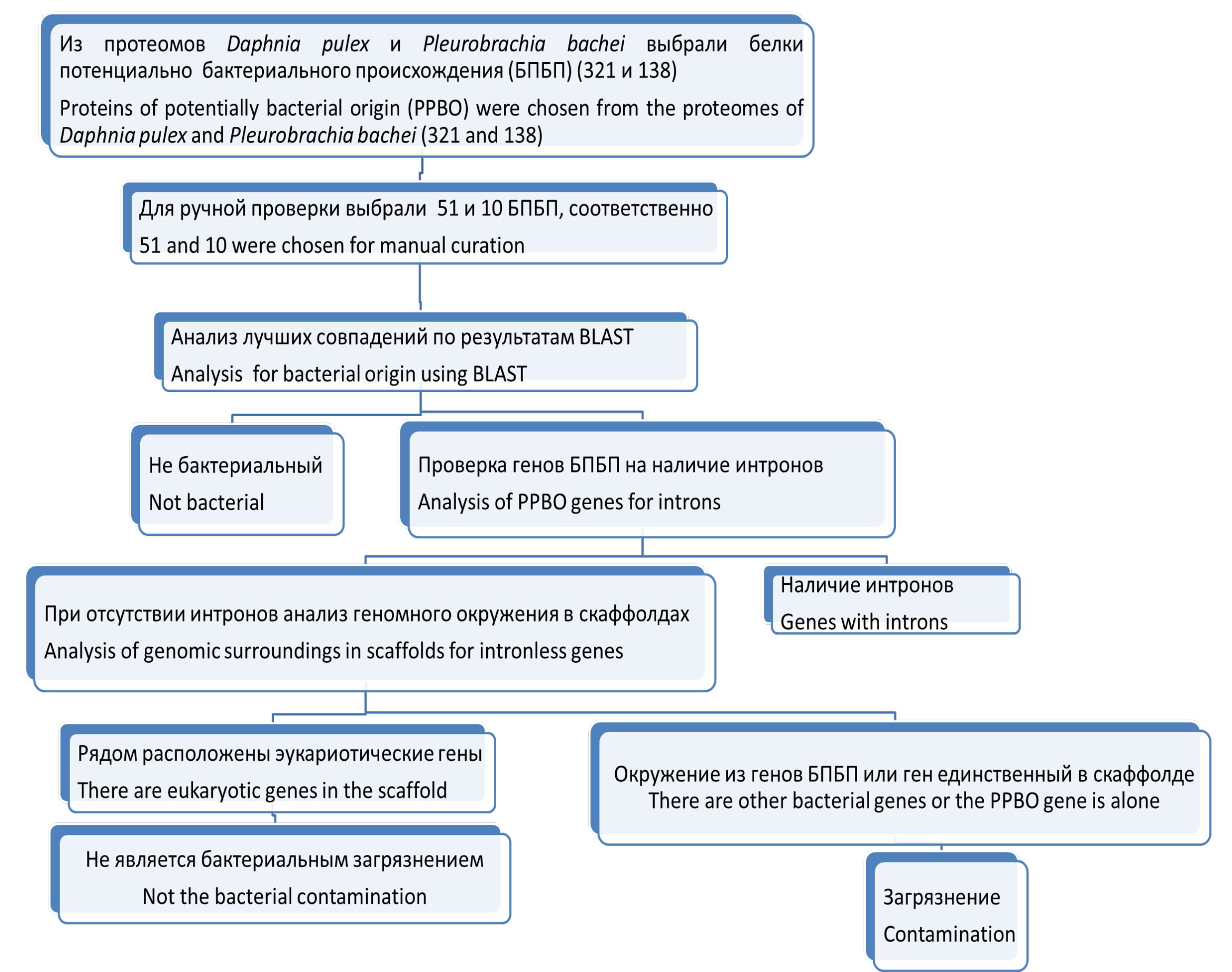


Рис. 3 Методика исследования БПБП  
Fig. 3 The analysis of PPBOs

Лучшее совпадение по BLAST Best BLAST-hit	Ручная проверка Manual curation		Автоматическая проверка Automatic analysis	
	Белки Proteins (51)	Скаффолды Scaffolds (36)	Белки Proteins (321)	Скаффолды Scaffolds (199)
Acidovorax	10	6	24	11
Comamonadaceae	16	11	41	24
Burkholderiales	22	16	72	38
Pseudomonas	15	9	25	13
др. бактерии other bacteria	12	7	49	49
не бактериальный not bacterial	2	4	224	99

Таблица 2. Статистика анализа БПБП и их скаффолдов  
Table 2. Statistics of analysis of PPBO and their scaffolds

## Results

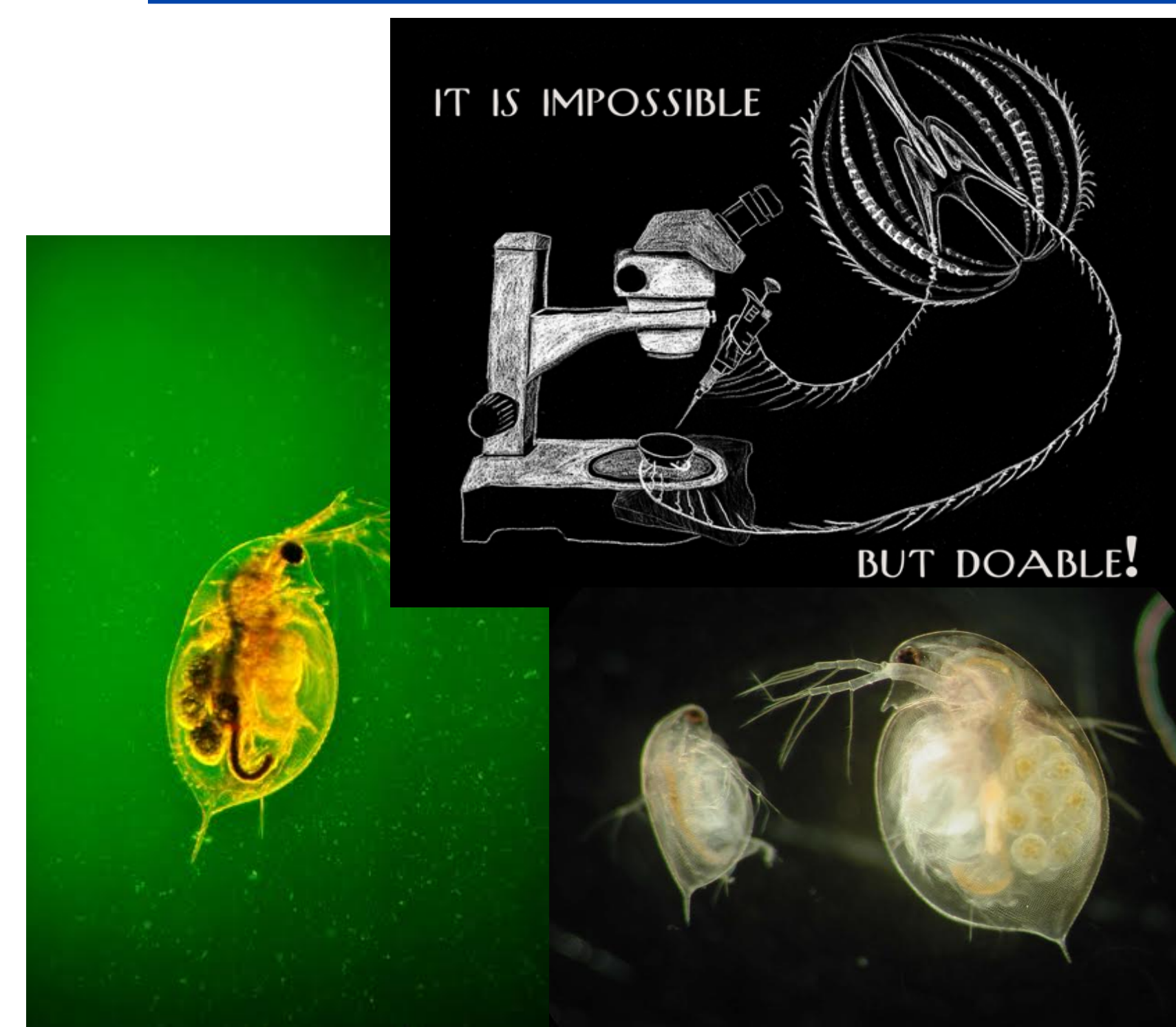


Рис. 2 (Fig. 2) *Pleurobrachia bachei*, *Daphnia Pulex*

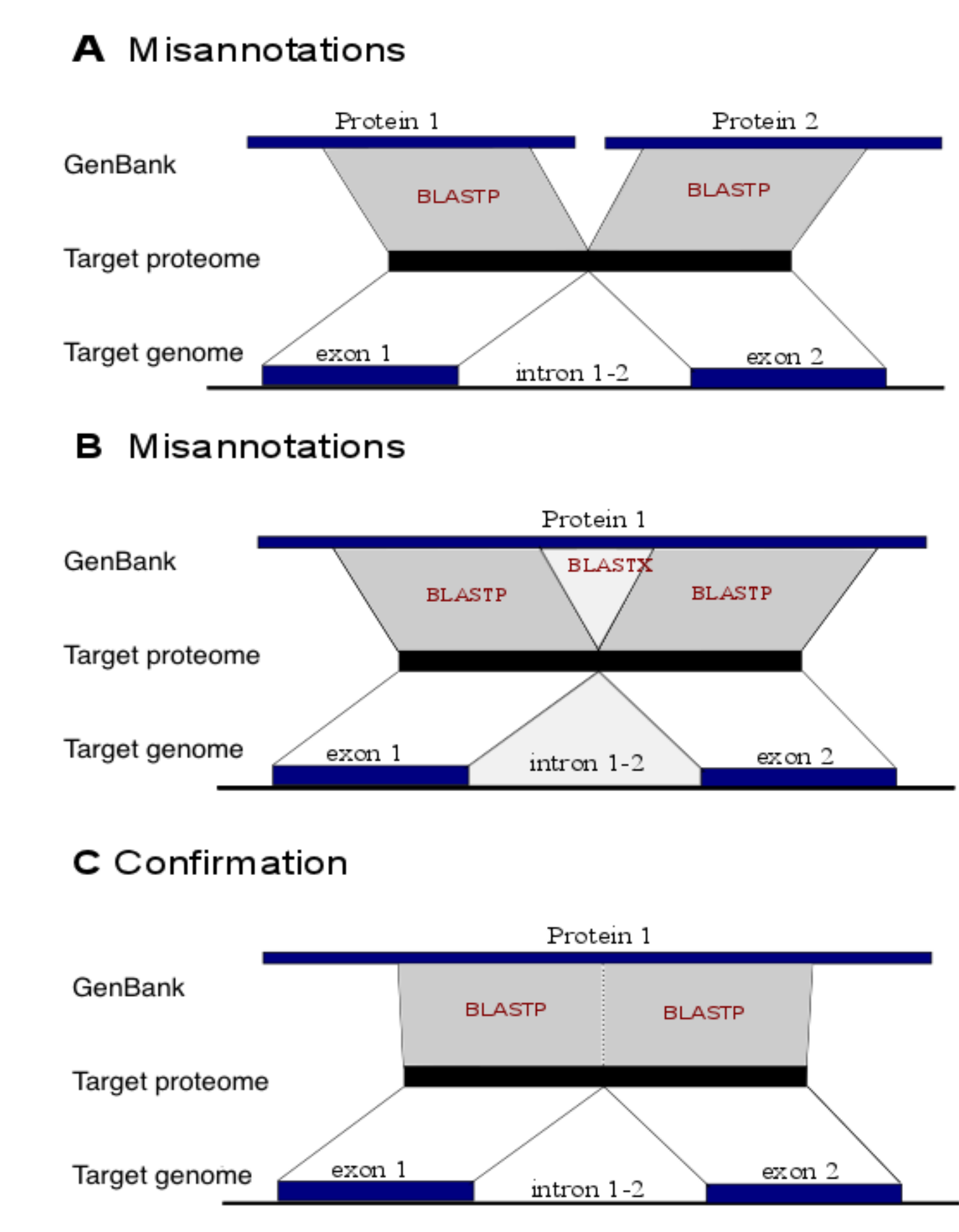


Рис. 4 Методика проверки интронов  
Fig. 4 The intron validation procedure

	Всего генов Total number of genes	Подозрительных на б.к. происхождение Potential Bacterial Origin	Выбрано для ручной проверки Selected for manual curation	Классифицировано как бактериальные Genes classified as bacterial	Из них многоэкзонных Multieukonic	Интроны аннотированы неверно Introns are annotated incorrectly
Гребневик (Pleurobrachia bachei)	19524	138	10	6	3	0
Дафния (Daphnia pulex)	30611	321	51	49	10	8 из 10

Таблица 1. Сводная статистика предварительной проверки БПБП  
Table 1. Statistics of the preliminary manual curation

## Introduction

- Modern approaches to the process of sequencing are not free of problems. Among them:
  - Sequenced organisms may have of symbionts, whose genetic material is present in the experimental samples
  - There are other sources of contamination by genetic material of other organisms during the preparation of the samples
  - The length of DNA fragments sequenced in one read is limited (~70-1000 bp). As a result, the assembly of reads is required. Reads are assembled into contigs, and contigs are arranged into scaffolds. There is a probability of incorrect arrangement of reads
  - The process of sequencing produces numerous errors and some genome parts of known length remain unsequenced. These problems are due to technical features of the sequencing process and/or due to DNA regions which are difficult for sequencing
  - The procedure for distinguishing bacterial contamination was established for incompletely assembled genomes of eukaryotic organisms. It allowed to describe the endosymbiont of *Hydra magnipapillata* and to confirm the existence of bacterial endosymbionts of *Nematostella vectensis*

## Aims

- Our aim is to check the recently sequenced genomes of *Pleurobrachia bachei* and *Daphnia pulex* for bacterial contamination and make a hypothesis about its origin (fig. 2).

## Conclusions

- In *Pleurobrachia bachei*, we failed to find bacterial contamination. The analyzed genes of bacterial origin either had introns or were merged into scaffolds with eukaryotic genes. It indicates that they were transferred to the genome horizontally.
- In *Daphnia pulex* we found evidence of bacterial contamination by several bacterial species. They are not yet described species from the family Comamonadaceae, Burkholderiales, Betaproteobacteria and not yet described bacteria from the genus *Pseudomonas* (Gammaproteobacteria). We suggested that these bacteriae are symbionts of *Daphnia pulex*.

## Materials and Methods

- The analysis of PPBO from *Pleurobrachia bachei* and *Daphnia pulex* is illustrated by the chart in fig. 3.
- The validation of introns for genes, coding PPBO, is illustrated by the scheme in fig. 4.