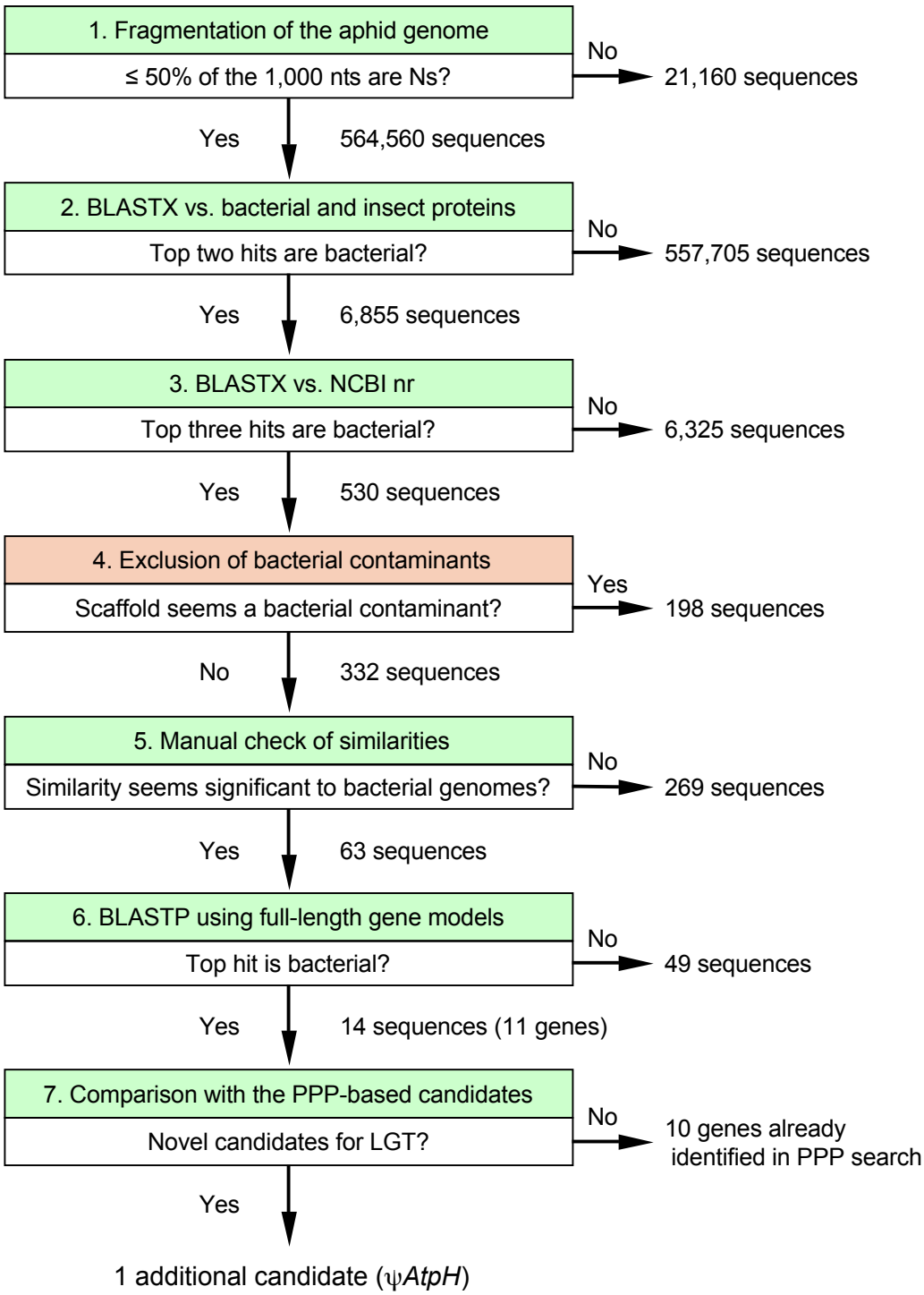
1. Fragmentation of the aphid genome
≤ 50% of the 1,000 nts are Ns?
No
21,160 sequences
Yes
564,560 sequences
2. BLASTX vs. bacterial and insect proteins
Top two hits are bacterial?
No
557,705 sequences
Yes
6,855 sequences
3. BLASTX vs. NCBI nr
Top three hits are bacterial?
No
6,325 sequences
Yes
530 sequences
4. Exclusion of bacterial contaminants
Scaffold seems a bacterial contaminant?
Yes
198 sequences
No
332 sequences
5. Manual check of similarities
Similarity seems significant to bacterial genomes?
No
269 sequences
Yes
63 sequences
6. BLASTP using full-length gene models
Top hit is bacterial?
No
49 sequences
Yes
14 sequences (11 genes)
7. Comparison with the PPP-based candidates
Novel candidates for LGT?
No
10 genes already identified in PPP search
Yes
1 additional candidate (ψ*AtpH*)

Figure S2