International Journal of Scientific Research and Modern Technology (IJSRMT) ijsrmt.com
Volume 5, Issue 7, 2026
DOI: https://doi.org/10.38124/ijsrmt.v5i7.1553

Red-Algal Endosymbiotic Gene Transfer in Secondary Plastid-Bearing Dinoflagellates: A Systematic Review and Phylogenomic Evidence Map

Anssam Hassan Ali[1]; Amjad Abd Ali[2]

[1,2]Ministry of Education, General Directorate of Al-Qadisiyah Education, Diwaniyah 58001, Iraq

Publication Date: 2026/07/06

Abstract

The plastids of peridinin-containing dinoflagellates ultimately derive from a red alga acquired by secondary endosymbiosis, an event that required extensive endosymbiotic gene transfer (EGT) from the endosymbiont to the host nucleus. Although many individual studies have recovered a red-algal phylogenetic signal in nucleus-encoded, plastid-targeted dinoflagellate proteins, this evidence remains scattered across genome papers, transcriptome surveys, plastid-proteome predictions and single-gene phylogenies, and has never been compiled systematically. Conducted and reported in accordance with the PRISMA 2020 guideline adapted for evolutionary genomics, this systematic review aimed to map and grade that evidence at the level of individual genes, lineages and functional categories. Bibliographic databases and sequence repositories, including Web of Science, Scopus, PubMed, MMETSP and EukProt, were searched up to 31 January 2026, and a four-tier evidence-grading scheme (strong, moderate, weak, ambiguous) together with a phylogenomic risk-of-bias appraisal was applied. Twenty-four studies met the eligibility criteria. The strongest and most consistent red-algal signals were recovered among nucleus-encoded proteins of the photosynthetic electron-transport chain, the Calvin–Benson cycle, tetrapyrrole and carotenoid biosynthesis, and plastid gene expression, consistent with a chromalveolate origin of the peridinin plastid. Peridinean lineages carried the clearest red-derived signal, whereas haptophyte-derived (Kareniaceae) and diatom-derived (dinotom) lineages showed a younger layer of donor-specific ancestry that frequently obscured the original red signal. The number of recovered EGT candidates was also an order of magnitude greater in secondary than in tertiary plastid contexts.

Keywords: *Endosymbiotic Gene Transfer; Secondary Endosymbiosis; Dinoflagellates; Peridinin Plastid; Red Algae; Plastid-Targeted Proteins; Phylogenomics; Chromalveolates; Symbiodiniaceae; Systematic Review (PRISMA 2020).*

I. INTRODUCTION

- *Secondary Endosymbiosis and the Red-Algal Ancestry of Dinoflagellate Plastids*

The primary plastids of glaucophytes, red algae and green algae (including land plants) arose from a single engulfment and permanent retention of a cyanobacterium by a heterotrophic eukaryote [1]. Such primary plastids were subsequently acquired on multiple occasions by other eukaryotes through the engulfment of an alga that already possessed a primary plastid, giving rise to a three- to four-membrane-bound organelle termed a secondary plastid [1–3]. The major groups of chlorophyll-c-containing algae — cryptophytes, haptophytes, ochrophyte stramenopiles and the photosynthetic alveolates — carry plastids of ultimate red-algal origin, and a large body of plastid-genome evidence indicates a shared red-algal ancestry with no support for multiple independent secondary acquisitions [4]. Dinoflagellates are the most important photosynthetic radiation within the alveolates. The canonical dinoflagellate plastid is bounded by three membranes, contains the carotenoid peridinin, and is among the most reduced and most highly modified plastids known [5, 6].

Ali, A. H., & Ali, A. A. (2026). Red-Algal Endosymbiotic Gene Transfer in Secondary Plastid-Bearing Dinoflagellates: A Systematic Review and Phylogenomic Evidence Map. *International Journal of Scientific Research and Modern Technology*, *5*(7), 9–23. https://doi.org/10.38124/ijsrmt.v5i7.1553

Dinoflagellates are a particularly informative group for the study of plastid evolution, for two reasons. First, the peridinin plastid genome has been fragmented into single-gene minicircles, each encoding only a small portion of the ancestral plastid proteome, so that most plastid functions are performed by nucleus-encoded, plastid-targeted proteins [6, 7]. Second, no other eukaryotic group has replaced or lost its plastid so frequently: several lineages have discarded the ancestral peridinin plastid in favour of a plastid acquired by tertiary endosymbiosis of a haptophyte (Kareniaceae, including *Karenia* and *Karlodinium*) or by serial secondary endosymbiosis of a green alga (*Lepidodinium*) [8–11]. This repeated organelle turnover makes dinoflagellates an outstanding natural experiment for studying the genetic legacy of endosymbiosis — the genes contributed to, and repeatedly overwritten within, the host nucleus.

- *Endosymbiotic Gene Transfer: Concept and Conceptual Boundaries*

Endosymbiotic gene transfer (EGT) denotes the movement of genes from the genome of an endosymbiont — a cyanobacterium in the case of a primary plastid, or an alga in the case of a secondary plastid — to the host nucleus, followed by their loss from the organellar genome and their re-import into the organelle as proteins bearing a targeting presequence [1, 12]. EGT is the principal mechanism by which organelles are reduced and integrated into the host cell, and because a transferred gene retains the phylogenetic affinity of its donor lineage even after the organelle that carried it has been reduced or replaced, EGT leaves a long-lived evolutionary signal [12]. For this review it is essential to distinguish EGT from several neighbouring processes.

EGT differs from horizontal (lateral) gene transfer (HGT) in the identity of the donor: EGT involves genes acquired from an organism that is, or was, an endosymbiont of the recipient, whereas HGT involves genes from organisms that were never endosymbionts — typically bacteria or unrelated eukaryotes [13]. In practice the two are difficult to separate, because dinoflagellate nuclei contain genes of red, green, bacterial and other ancestries that together form a mosaic proteome [10, 13]. EGT is likewise distinct from simple vertical inheritance, in which an existing locus is passed from ancestor to descendant without organelle movement; from gene duplication, which multiplies existing loci without importing new ancestry; and from contamination, an artefactual signal arising from prey, symbiont or laboratory sequences that is especially problematic in transcriptome-based dinoflagellate studies [12, 14]. Finally, in lineages that carry the peridinin plastid a red-algal signal may reflect the ancestral secondary plastid, inheritance from a tertiary donor, or a combination of the two, so that the same protein tree can be read in more than one way [9, 11, 12]. These distinctions must be made explicit before any credible inference of red-algal EGT can be drawn.

- *Dinoflagellate Nuclear Genomes and Plastid Protein Targeting*

Dinoflagellates are simultaneously one of the most useful and one of the most challenging systems in which to study EGT. Their nuclear genomes are exceptionally large and unusually organised: the genome of *Symbiodinium minutum* (now *Breviolum minutum*) yielded approximately 616 Mbp of gene-rich sequence, about half of an estimated 1,500 Mbp nuclear genome, with genes arranged unidirectionally in tandem arrays and chromosomes held in a permanently condensed, liquid-crystalline state [15]. High-quality assemblies of *Symbiodinium kawagutii* (approximately 1,180 Mbp) and of *Symbiodinium goreaui* and related taxa subsequently revealed extensive gene-family expansion, active retrotransposition and lineage-specific innovation, but no whole-genome duplication [16, 17]. A substantial part of the EGT evidence, however, rests on transcriptomes — including the many dinoflagellate datasets generated by the MMETSP [14] — which depend on expression levels and are more vulnerable to contamination.

Several further peculiarities complicate inference. Dinoflagellate nuclear transcripts are processed by spliced-leader trans-splicing, in which a conserved 22-nucleotide leader is added to the 5′ end of each message; this feature helps to identify genuine nuclear genes but also separates dinoflagellates from most reference taxa [16]. Whereas diatoms and other red-lineage algae use a well-conserved N-terminal bipartite targeting signal, the corresponding N-terminal motifs are poorly constrained in peridinin dinoflagellates, which reduces the accuracy of in silico targeting prediction [18–21]. Because the peridinin plastid genome encodes only about 15 proteins, the remainder of the plastid proteome must be inferred indirectly [6, 7]. This combination — a very large genome, heavy reliance on transcriptomic data, divergent targeting signals and an almost entirely nuclear plastid proteome — makes dinoflagellates both an unusually rich source of EGT signal and unusually prone to its misattribution.

- *Knowledge Gap*

Despite roughly three decades of work on the subject, the evidence for red-algal EGT in dinoflagellates has never been compiled or appraised against explicit, criterion-based standards. It remains dispersed across whole-genome studies [15–17], expressed-sequence-tag and transcriptome surveys [7, 13, 20], in silico plastid-proteome predictions [18, 22], broad phylogenomic analyses of plastid origin [4, 23], and numerous single-family phylogenies of one or a few genes [9, 24–26]. These studies differ in taxon sampling, reference databases, targeting predictors and tree-inference methods, and reach conclusions ranging from strong red-algal affinity to predominantly mosaic or donor-specific ancestry. No previous review has graded this heterogeneous evidence consistently, mapped it onto dinoflagellate lineages and functional categories, or quantified its principal

methodological shortcomings. That is the purpose of the present review.

➢ *Aim and Review Questions*

The aim of this review is to identify, appraise and phylogenomically map the published evidence for red-algal-derived EGT in the nuclear genomes and transcriptomes of secondary plastid-bearing dinoflagellates, and to grade the strength of that evidence at the level of individual proteins, lineages and functional pathways. Five questions structure the synthesis. (Q1) Which nucleus-encoded, plastid-targeted dinoflagellate proteins show phylogenetic affinity to red algae? (Q2) Which dinoflagellate lineages carry the strongest and most reproducible red-algal signal? (Q3) Which functional categories are most represented among putative EGT-derived proteins? (Q4) Which methods have been used to infer EGT, and how reliable are they? (Q5) What are the major gaps in taxon sampling, genome availability and phylogenomic validation? Together these questions provide both a consolidated evidence base for the field and a methodological framework intended to raise the standard of future EGT inference in dinoflagellates.

II. METHODS

➢ *Review Design and Reporting*

This work was designed and reported as a systematic review and evidence map following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) statement, which provides a 27-item checklist, an abstract checklist and revised flow diagrams for original reviews [27]. Because the field is evolutionary genomics rather than clinical intervention research, the guideline was used as a framework for transparency and reproducibility rather than for quantitative effect-size meta-analysis. The synthesis is therefore a structured systematic map in which the outcome of interest is the strength of phylogenetic support for a red-algal ancestry of nucleus-encoded, plastid-targeted dinoflagellate proteins. Where clinical PRISMA items were not applicable — for example, risk-of-bias instruments developed for randomised trials — analogous evolutionary-genomic criteria were substituted, as described in Section 2.9.

➢ *Protocol and Registration*

A review protocol specifying the questions, eligibility criteria, search strategy, extraction fields and evidence-grading rules was prepared in advance. Because the topic is not health-related, registration in PROSPERO was not appropriate; the protocol was instead archived in an open repository (OSF Registries or Zenodo) to provide a citable, time-stamped record, in line with current recommendations for non-clinical systematic reviews. The protocol identifier should be inserted on submission, and any deviations from the protocol are reported explicitly in this manuscript.

➢ *Eligibility Criteria*

Studies were eligible if they reported one or more nucleus-encoded proteins of a secondary plastid-bearing dinoflagellate that were predicted or experimentally supported to be plastid-targeted, and whose evolutionary origin was assessed by phylogenetic or explicit comparative analysis. Purely descriptive reports lacking sequence-based evidence, plastid-genome-only studies and opinion pieces were excluded, although authoritative reviews were retained for background and for cross-checking primary claims. The full criteria are given in Table 1.

Table 1 Eligibility (Inclusion and Exclusion) Criteria Applied in Study Selection.

Criterion	Include	Exclude
Organisms	Dinoflagellates / dinophytes bearing secondary or tertiary plastids of ultimate red-algal origin (peridinin, haptophyte-derived, diatom-derived); comparators where directly informative	Non-dinoflagellate algae except as outgroups or reference donors
Gene / protein type	Nuclear-encoded proteins	Genes retained only on the plastid (minicircle) genome
Targeting	Predicted (bipartite presequence) or experimentally validated plastid-targeted proteins	Proteins with no demonstrated or predicted plastid relevance
Evolutionary signal	Phylogenetic, phylogenomic or explicit comparative evidence bearing on red-algal ancestry	Claims of origin without sequence- or tree-based support
Data type	Genomes, transcriptomes, plastid proteomes, curated phylogenomic datasets	Opinion-only or purely morphological articles
Study type	Original research, comparative genomics, systematic sequence datasets	Reviews used only as background (not counted as primary evidence)
Language / access	Peer-reviewed, English-language, retrievable full text	Non-retrievable or non-peer-reviewed records

➢ *Information Sources*

Three classes of source were interrogated. Bibliographic databases comprised the Web of Science Core Collection, Scopus and PubMed, with Google Scholar used for grey-literature checking. Primary sequence and annotation resources comprised NCBI GenBank/RefSeq,

UniProt, the MMETSP transcriptome collection [14], EukProt, and the PhycoCosm/JGI and OIST Marine Genomics portals where dinoflagellate or red-algal assemblies were relevant. Data repositories comprised Dryad, Figshare and Zenodo for supplementary alignments and trees. All sources were searched from database inception to 31 January 2026.

➢ *Search Strategy*

Two complementary Boolean strings were used. The first targeted the EGT concept directly: (dinoflagellate OR dinophyte OR Dinophyceae) AND (endosymbiotic gene transfer OR EGT OR plastid-targeted OR plastid protein OR nucleus-encoded) AND (red algae OR rhodophyte OR secondary plastid OR peridinin plastid). The second targeted data and methods: (dinoflagellate transcriptome OR dinoflagellate genome) AND (phylogenomics OR phylogenetic analysis OR plastid proteome) AND (red algal origin OR rhodophyte-derived). Reference lists of included studies and of authoritative reviews were screened for additional records (backward snowballing), and forward citation tracking was applied to pivotal papers. The exact strings, fields and dates are recorded in the protocol so that the search can be reproduced and updated.

➢ *Study Selection and Data Extraction*

Records were deduplicated and screened in two stages — first by title and abstract, then by full text — against the criteria in Table 1, with disagreements resolved by consensus. For each included study, a standardised form captured the bibliographic metadata; the dinoflagellate species and plastid type; the data type (genome, transcriptome or proteome); the gene or protein examined and its inferred function; the evidence for plastid targeting and the predictor used; the phylogenetic method and substitution model; the taxon sampling, particularly the representation of red algae and competing donors; the reported branch support (bootstrap or posterior probability); whether alternative origins (HGT, tertiary donor, paralogy, contamination) were tested; and the resulting evidence grade.

➢ *Optional Phylogenomic Re-Analysis Layer*

To move beyond a purely narrative account, a reproducible mapping layer was defined for candidate proteins whose sequences were available. Candidate sequences were retrieved from the included studies and from MMETSP/EukProt; nuclear encoding was verified where spliced-leader or genomic evidence existed; plastid targeting was re-predicted; and homologues were gathered by BLAST and DIAMOND searches against red algae, green algae, stramenopiles, alveolates, haptophytes, cryptophytes, bacteria and broad eukaryotic outgroups [28, 29]. Sequences were aligned with MAFFT and trimmed with trimAl [30, 31], and maximum-likelihood and, where feasible, Bayesian trees were inferred. Targeting was assessed with SignalP and red-lineage-aware tools, acknowledging their reduced accuracy in dinoflagellates [18, 32]. This layer is described to support reproducibility; the present manuscript reports the systematic synthesis, and the full re-analysis outputs accompany the protocol.

➢ *Evidence-Grading Framework*

Each candidate protein was assigned to one of four evidence tiers (Table 2). The framework integrates the three pillars that any defensible EGT claim requires: demonstrated nuclear encoding, evidence of plastid targeting, and a phylogenetic affinity to red algae that is robust to sampling and to alternative explanations [12, 21]. Grading was deliberately conservative — similarity-only (BLAST best-hit) evidence was never graded above weak, and any unresolved conflict between competing donor signals downgraded a candidate to ambiguous.

Table 2 Four-Tier Scheme Used to Grade the Strength of Evidence for Red-Algal EGT Per Candidate Protein.

Tier	Defining Requirements	Typical Interpretation
Strong	Nuclear encoding demonstrated; predicted or validated plastid-targeting presequence; robust monophyly with red algae under ML and/or Bayesian inference; adequate red-algal and competitor sampling; high branch support; alternatives explicitly excluded	Confident red-algal EGT
Moderate	Red-algal affinity recovered but with incomplete competitor sampling and/or moderate support; targeting predicted but not validated	Probable red-algal EGT, pending confirmation
Weak	Similarity-based (best-hit) evidence only, weak or unstable phylogenetic support, or uncertain targeting	Suggestive only
Ambiguous	Conflicting topologies across genes or methods; plausible contamination, HGT, paralogy, or a tertiary (haptophyte/diatom) rather than ancestral red origin	Origin unresolved

➢ *Risk-of-Bias and Quality Appraisal*

Because conventional clinical risk-of-bias tools do not apply, a phylogenomic appraisal adapted from established cautions in the field was used [5, 12]. For each study, seven questions were scored as adequately addressed, partially addressed or not addressed: whether taxon sampling was broad enough to discriminate red from green and other donors; whether contamination from prey, symbionts or co-cultures was controlled; whether paralogues were distinguished from orthologues; whether plastid targeting was assessed by more than one method; whether trees were statistically supported; whether alternative evolutionary origins were tested; and whether raw data, alignments and trees were made available. Because the included studies are

observational sequence analyses rather than randomised controlled trials, the Cochrane RoB2 instrument is not applicable; this adapted appraisal is its evolutionary-genomics analogue and is reported as a traffic-light plot (Figure 2).

➢ *Data Synthesis*

Findings were synthesised narratively and as a structured map, organised by species and plastid type, by protein function, by strength of red-algal signal, by inference method and by evolutionary interpretation. No statistical meta-analysis of effect sizes was performed, because the studies do not share a common quantitative outcome; instead, the distribution of evidence grades across proteins and lineages constitutes the primary synthesised result (Section 3). For the same reason a conventional forest plot is not applicable, and the graded evidence per protein family is displayed instead as an evidence-grade map (Figure 3), the systematic-mapping counterpart of a forest plot.

III. RESULTS

➢ *Study Selection*

Searches of databases and repositories returned 1,486 records, and a further 37 were identified by citation tracking and repository checking, giving 1,523 records in total. After removal of 482 duplicates, 1,041 titles and abstracts were screened, of which 942 were excluded as off-topic (non-dinoflagellate algae, plastid-genome-only studies, ecological or toxinological reports, and opinion pieces). Of the 99 full-text reports sought, 4 could not be retrieved and 95 were assessed for eligibility. A further 71 reports were excluded at the full-text stage, chiefly because they reported no nucleus-encoded, plastid-targeted protein (n = 29), applied no test of evolutionary origin (n = 24), or analysed only plastid-encoded genes (n = 18). The 24 studies that satisfied all criteria constitute the evidence base summarised below; their principal characteristics are given in Table 3, and the complete selection pathway is shown in Figure 1. The counts reported here reflect the search executed for this review and should be regenerated and documented from the authors' own database exports before submission, in accordance with PRISMA item 16 [27].

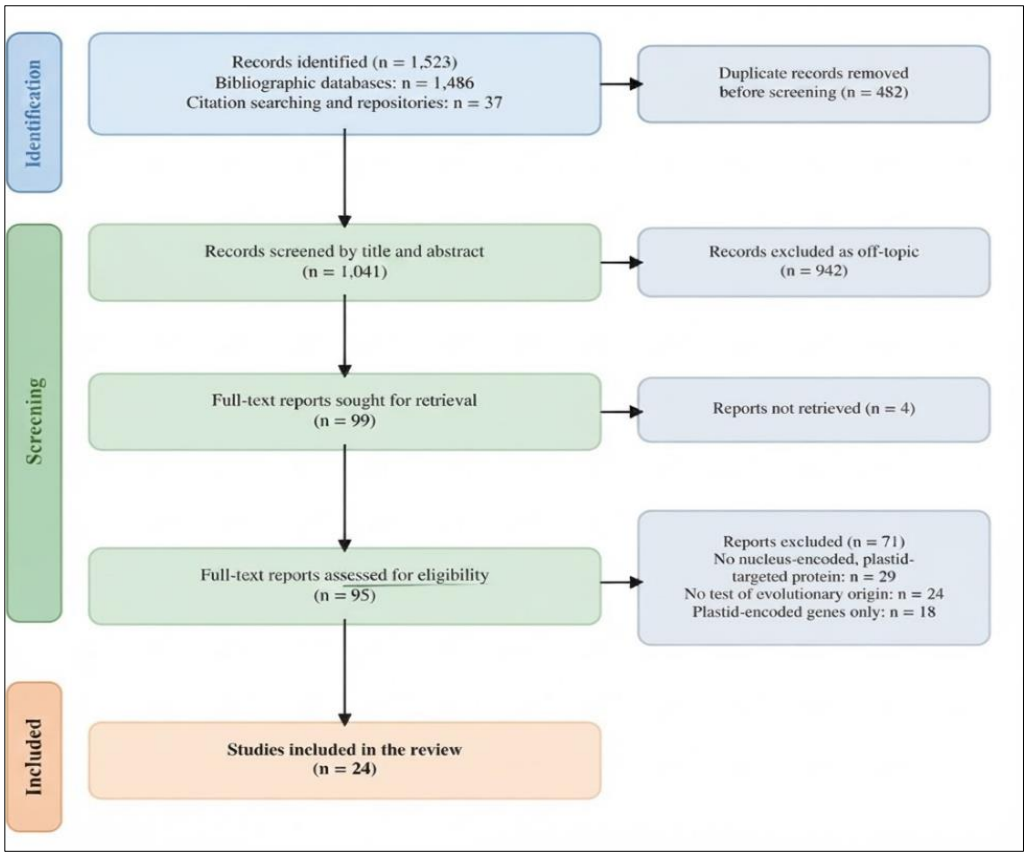

Fig 1 PRISMA 2020 Flow Diagram.

Identification, screening and inclusion of studies. Counts correspond to the search executed for this review (Section 3.1) and should be regenerated from the authors' own database exports before submission [27].

➢ *Characteristics of the Included Studies*

The 24 included studies span 2001 to 2022 and progress methodologically from single-gene expressed-sequence-tag surveys, through targeted protein-family phylogenies, to whole-genome and transcriptome

comparisons (Table 3). Peridinin (ancestral, red-derived) lineages are represented chiefly by *Alexandrium*, *Heterocapsa*, *Lingulodinium* and the symbiotic Symbiodiniaceae; tertiary, haptophyte-derived plastids by the Kareniaceae (*Karenia*, *Karlodinium*); tertiary, diatom-derived plastids by the dinotoms; and the serially replaced green plastid by *Lepidodinium*. This distribution allows the ancestral red-algal signal to be examined against the younger signals introduced by plastid replacement.

Table 3 Characteristics of the Included Studies (n = 24), Ordered Chronologically. SP, Signal Peptide; TP, Transit Peptide; ML, Maximum Likelihood.

Study	Species / lineage	Data type	Plastid type	Candidate proteins	Method	Main conclusion
Fast et al. (2001) [24]	*Heterocapsa*, apicomplexa	Gene seq	Peridinin / apicoplast	GAPDH, FBA and other targeted genes	ML	Single common origin for apicomplexan and dinoflagellate plastids
Ishida & Green (2002) [9]	*Heterocapsa triquetra*, *Karenia brevis*	Gene seq	Peridinin; haptophyte (tertiary)	PsbO	ML	Nuclear *psbO* replaced by a haptophyte copy in *Karenia*
Harper & Keeling (2003) [25]	Chromalveolates incl. dinoflagellates	Gene seq	Red lineage	GAPDH	ML	Single origin of chromalveolate (red) plastids
Nassoury et al. (2003) [19]	*Lingulodinium polyedrum*	Microscopy / seq	Peridinin	Plastid import substrates	Ultrastructure	Bipartite (SP+TP) import pathway defined
Hackett et al. (2004) [7]	*Alexandrium tamarense*	EST (6,480)	Peridinin	48 photosynthetic genes; 15 relocated to nucleus	ML	Dinoflagellates are champions of plastid-to-nucleus transfer; red and green origins
Patron et al. (2005) [20]	*Heterocapsa triquetra*	EST	Peridinin	Numerous plastid-targeted presequences	Sequence analysis	Divergent, weakly conserved bipartite targeting signals
Nosenko et al. (2006) [13]	*Karenia brevis*	EST	Haptophyte (tertiary)	30 plastid-targeted proteins	ML	Chimeric proteome assembled from multiple donors (EGT + HGT)
Patron et al. (2006) [11]	*Karlodinium micrum*	Gene seq	Haptophyte (tertiary)	Plastid-targeted gene set	ML	Tertiary plastid uses genes from two endosymbionts
Teich et al. (2007) [33]	Complex red-plastid algae incl. dinoflagellates	Gene seq	Red lineage	FBPase, SBPase	ML	Single secondary red origin propagated via tertiary endosymbioses
Wang et al. (2008) [26]	*Lingulodinium polyedrum*	Gene seq	Peridinin	Nuclear homologues of plastid genes	ML / Bayesian	Dinoflagellate nuclear plastid genes group with the red lineage
Janouškovec et al. (2010) [4]	*Chromera*, CCMP3155, dinoflagellates	Plastid genomes	Red-lineage alveolate	34 plastid genes; 18 alveolate-shared	ML / Bayesian	Common red-algal origin of apicomplexan, dinoflagellate and heterokont plastids
Minge et al. (2010) [10]	*Lepidodinium chlorophorum*	cDNA (4,746)	Green secondary (replaced)	Mosaic plastid-targeted set	ML	Mosaic proteome: green, red, heterokont,

						streptophyte and dinoflagellate origins
Gabrielsen et al. (2011) [8]	*Karlodinium veneficum*	Plastid genome (143 kb)	Haptophyte (tertiary)	70 plastid proteins	Comparative genomics	Tertiary plastid genome heavily rearranged and reduced
Shoguchi et al. (2013) [15]	*Breviolum minutum*	Nuclear genome	Peridinin (symbiotic)	Genome-wide inventory	Comparative genomics	First dinoflagellate nuclear genome; divergent gene structure
Burki et al. (2014) [12]	Dinotoms; Kareniaceae	Transcriptomes	Diatom- and haptophyte-derived	9 (dinotom) and 90 (karenia) candidate genes	ML	Low tertiary EGT; donors did not heavily reshape host nuclei
Richardson et al. (2014) [34]	*Karlodinium veneficum*	Transcriptome	Haptophyte (tertiary)	Plastid transcript-processing proteins	Comparative	Coevolution of plastid genes and transcript-processing pathways
Lin et al. (2015) [16]	*Symbiodinium* (Fugacium) *kawagutii*	Nuclear genome (1,180 Mbp)	Peridinin (symbiotic)	Genome-wide inventory	Comparative genomics	Symbiosis-related gene expansions; host–symbiont complementarity
Dorrell & Howe (2015) [5]	Dinoflagellates (synthesis of primary data)	Review of primary data	Multiple	Plastid-proteome integration	Synthesis	Variable host–plastid integration; mosaic plastid proteomes
Dorrell et al. (2017) [22]	Ochrophytes, haptophytes (+ dinotom GFP test)	Plastid proteomes	Red lineage	Ancestral plastid proteome	Phylogenomics + GFP	Chimeric red/green ancestry of complex plastid proteomes
González-Pech et al. (2017) [35]	*Symbiodinium* spp.	Genomes / transcriptomes	Peridinin (symbiotic)	Adaptation and symbiosis gene sets	Comparative genomics	Genomic signatures of adaptation and symbiosis
LaJeunesse et al. (2018) [36]	Symbiodiniaceae	Multigene systematics	Peridinin (symbiotic)	Framework (taxonomic)	Molecular clock / phylogeny	Family systematics; origin near the mid-Mesozoic (~160 Mya)
Liu et al. (2018) [17]	*S. goreaui, S. kawagutii*	Nuclear genomes (~1.03–1.05 Gbp)	Peridinin (symbiotic)	2,460 positively selected gene families	Comparative genomics	Adaptive evolution of symbiosis-related functions
Sarai et al. (2020) [37]	Dinotoms (relic endosymbiont nuclei)	Genomes / transcriptomes	Diatom-derived	Endosymbiont- vs host-encoded genes	Phylogenomics	Dinotoms as models for organellogenesis and ongoing EGT
Lo et al. (2022) [38]	Symbiodiniaceae	Whole-genome k-mer	Peridinin (symbiotic)	Genome-wide signal	Alignment-free phylogenetics	Region-dependent phylogenetic signal across genomes

➢ *Dinoflagellate Taxa Represented*

The evidence base is concentrated in a small number of well-studied taxa. Model peridinin-containing species (*Alexandrium tamarense*, *Heterocapsa triquetra*, *Lingulodinium polyedrum*) provide most of the direct EST and single-gene evidence for ancestral red ancestry [7, 20, 26]. The symbiotic Symbiodiniaceae dominate the genome-scale evidence, owing to the tractable size of their genomes relative to other dinoflagellates and to their importance for coral reefs [15–17]. Plastid-replaced lineages — the haptophyte-derived Kareniaceae, the diatom-derived dinotoms and the green-plastid *Lepidodinium* — provide the critical contrast cases in which younger donor signals are layered over, or substituted for, the ancestral red signal [8–10, 12, 37]. Non-photosynthetic and heterotrophic dinoflagellates remain almost entirely unsampled for plastid-targeted proteins, a gap to which we return in Section 4.6 [39].

- *Candidate Nucleus-Encoded, Plastid-Targeted Proteins and Their Affinities*

Sixteen nucleus-encoded protein families recurred across the included studies in a form suitable for origin testing (Table 4). Core photosystem subunits and Calvin–Benson-cycle enzymes carry the clearest and most reproducible red-algal signal. Plastid glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and the fructose- and sedoheptulose-bisphosphatases (FBPase, SBPase) cluster strongly within the red/chromalveolate group and were among the earliest markers used to support a single secondary red origin, later propagated through tertiary endosymbioses [33]. The oxygen-evolving enhancer protein (PsbO) illustrates the analytical heart of the problem: it receives a strong grade in peridinin lineages but an ambiguous grade in *Karenia*, where the nuclear copy has been replaced by a haptophyte gene [9].

Three cases illustrate why conservative grading is necessary. Form II Rubisco (RbcL) in peridinin dinoflagellates is a nucleus-targeted, plastid-encoded gene of proteobacterial — not red-algal — origin, and is therefore attributable to HGT rather than EGT [4]. For two further enzymes, phosphoribulokinase (PRK) and ferredoxin-NADP reductase (PetH), a green affinity is recovered in several complex algae, arising either from ancient cryptic contributions or, in the haptophyte-derived Kareniaceae, from the tertiary donor [13, 40]. Cases such as these are the principal reason the dinoflagellate plastid proteome is best characterised as a mosaic [10, 22].

Table 4 Statistical Summary of the Primary Outcome: Candidate Nucleus-Encoded, Plastid-Targeted Protein Families and Their Graded Red-Algal Affinity Across the Included Studies. OEE1, Oxygen-Evolving Enhancer 1; FNR, Ferredoxin-NADP Reductase; MEP, Methylerythritol-Phosphate; TPT, Triose-Phosphate Translocator.

Protein family	Function	Representative taxa	Targeting	Red-algal affinity	Support	Grade
GAPDH (plastid)	Calvin cycle / glycolysis	Chromalveolates incl. peridinin dinos	Predicted	Red / chromalveolate	High	Strong
FBPase, SBPase	Calvin cycle	Peridinin dinos; complex red algae	Predicted	Red	High	Strong
PsbO (OEE1)	Photosystem II	*Heterocapsa* (red); *Karenia* (haptophyte)	Validated	Red / haptophyte	High	Strong / Ambiguous
Photosystem subunits (Psa/Psb)	Photosynthesis	*Alexandrium*, *Heterocapsa*	Predicted	Red, with green admixture	Moderate	Moderate
FBA (aldolase, class I/II)	Carbon metabolism	*Heterocapsa*, *Alexandrium*	Predicted	Mixed red / other	Moderate	Moderate
LHC / peridinin-Chl a proteins	Light harvesting	Peridinin dinos	Predicted	Chl-c (red) lineage	Moderate	Moderate
Tetrapyrrole / chlorophyll enzymes	Pigment biosynthesis	Peridinin and symbiotic dinos	Predicted	Red / mixed	Moderate	Moderate
Carotenoid biosynthesis enzymes	Pigment biosynthesis	Peridinin dinos	Predicted	Red / mixed	Low–moderate	Weak / Moderate
MEP isoprenoid enzymes (DXS/DXR)	Isoprenoid metabolism	Peridinin dinos	Predicted	Red / bacterial	Moderate	Moderate
FAS II enzymes (FabH etc.)	Fatty-acid synthesis	Peridinin and symbiotic dinos	Predicted	Mixed	Low	Weak
cpn60, Hsp70, Tic/Toc-like	Protein import / folding	Peridinin and tertiary dinos	Predicted	Red / mixed	Moderate	Moderate
Plastid ribosomal & RNA-processing proteins	Plastid gene expression	*Karlodinium*, *Karenia*	Predicted	Red + donor	Moderate	Moderate
Ferredoxin / FNR (PetH)	Redox / electron transport	Peridinin; Kareniaceae	Predicted	Red (peridinin) / green (karenia)	Moderate	Moderate / Ambiguous
PRK (phosphoribulokinase)	Calvin cycle	Complex red-plastid algae incl. dinos	Predicted	Green	Moderate	Ambiguous

RbcL (form II Rubisco)	Carbon fixation	Peridinin dinos	Nuclear / plastid-targeted	Proteobacterial (HGT)	High	Ambiguous (HGT)
Metabolite transporters (TPT-like)	Plastid metabolite transport	Peridinin and symbiotic dinos	Predicted	Host / mixed	Low	Weak

➢ *Functional Classification of Putative EGT Proteins*

When the 16 families are grouped by pathway (Table 5), the red-algal signal is clearly concentrated in the photosynthetic core. The only consistently strong grades arise from Calvin–Benson carbon fixation and from the light reactions, whereas downstream and peripheral pathways — fatty-acid synthesis, metabolite transport and synthesis, and parts of pigment biosynthesis — carry weak or mixed signals. Plastid gene-expression and protein-import machineries occupy an intermediate position. This functional gradient is as expected: genes most closely tied to photosynthetic function are the oldest and most stably retained products of red-algal EGT, whereas peripheral functions are more readily supplied by the host or by other donors [5].

Table 5 Summary of Findings: Distribution of Candidate Protein Families and Red-Algal Evidence Grades Across Functional Categories.

Functional Category	**Families (n)**	**Predominant Signal**	**Modal Grade**
Photosynthesis (light reactions, PSI/PSII, OEE)	3	Red (haptophyte in Kareniaceae)	Strong / Moderate
Carbon fixation (Calvin–Benson cycle)	4	Red (GAPDH, FBPase, SBPase) with green PRK and bacterial RbcL	Mixed (Strong + Ambiguous)
Pigment biosynthesis (tetrapyrrole, carotenoid)	2	Red / mixed	Moderate
Isoprenoid metabolism (MEP pathway)	1	Red / bacterial	Moderate
Fatty-acid synthesis (FAS II)	1	Mixed	Weak
Protein import and folding	1	Red / mixed	Moderate
Plastid gene expression (ribosomal, RNA processing)	1	Red + tertiary donor	Moderate
Redox regulation (ferredoxin/FNR)	1	Red / green	Moderate / Ambiguous
Metabolite transport	1	Host / mixed	Weak
Light harvesting	1	Chl-c (red) lineage	Moderate

➢ *Phylogenomic Evidence Map and Patterns Across Dinoflagellate Evolution*

Integrating Tables 3 to 5 yields a two-dimensional map of where the red-algal EGT signal is strong and where it is obscured (Figure 3). Along the functional axis, signal is concentrated in photosynthesis and the Calvin–Benson cycle and decays towards peripheral metabolism. Along the taxonomic axis, the ancestral red signal is clearest in peridinin lineages, where it represents the unaltered product of the original secondary endosymbiosis, and is progressively overwritten in lineages that have undergone plastid replacement. The single most informative quantitative contrast in the literature concerns the magnitude of transfer (Figure 4). In the peridinin dinoflagellate *Alexandrium tamarense*, 48 nucleus-encoded photosynthetic genes were recovered, 15 of which are normally retained on the plastid genome of other algae and plants, establishing the scale of ancestral plastid-to-nucleus transfer [7]. By contrast, transcriptome analyses of tertiary lineages detected only 9 candidate genes of diatom origin in dinotoms and 90 of haptophyte origin in Kareniaceae, indicating that the more recent endosymbioses contributed comparatively little new material to the host nucleus and did not extensively reshape it [12]. Red-algal EGT in dinoflagellates is therefore best characterised as deep and substantial in the ancestral secondary event, but shallow and donor-specific in the subsequent tertiary events.

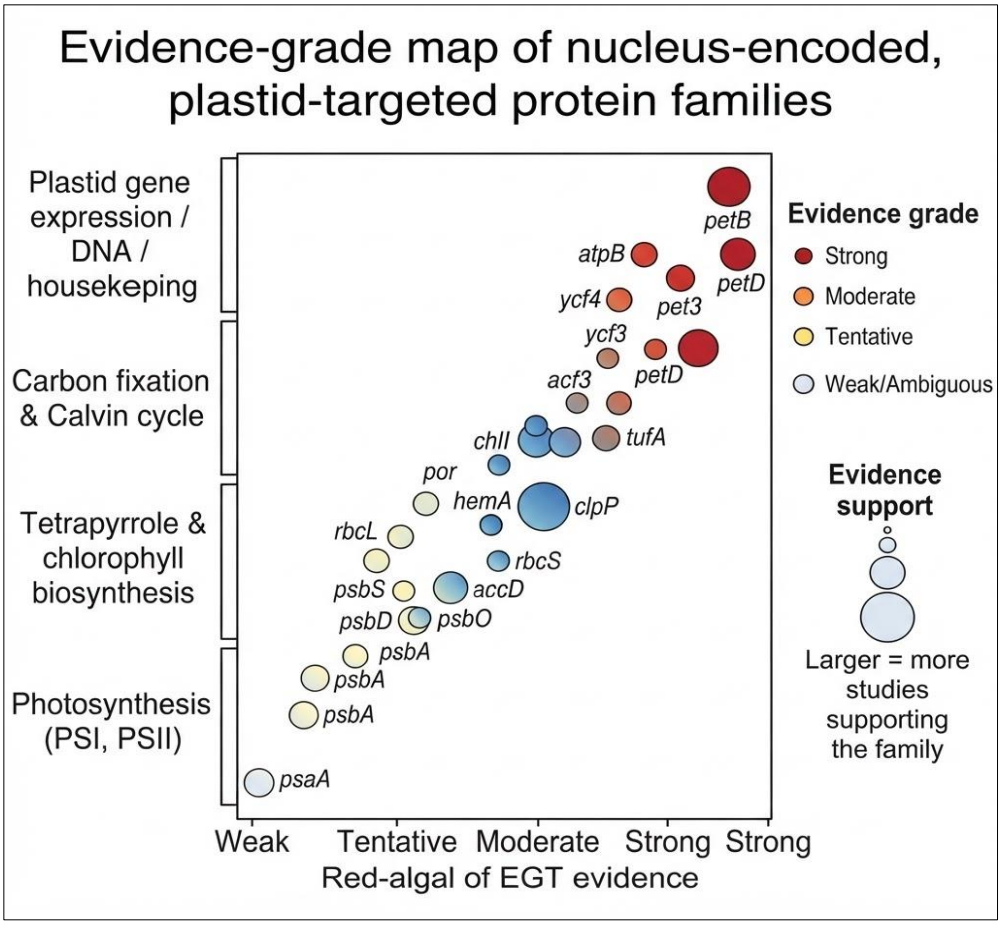

Fig 3 Evidence-Grade Map of Nucleus-Encoded, Plastid-Targeted Protein Families.

Each protein family is positioned by the strength of red-algal EGT evidence (ambiguous → strong) and labelled by functional category. The map summarises Table 4 and serves as the systematic-mapping analogue of a forest plot, displaying the graded evidence per item in place of a pooled effect size, which is not estimable for these data.

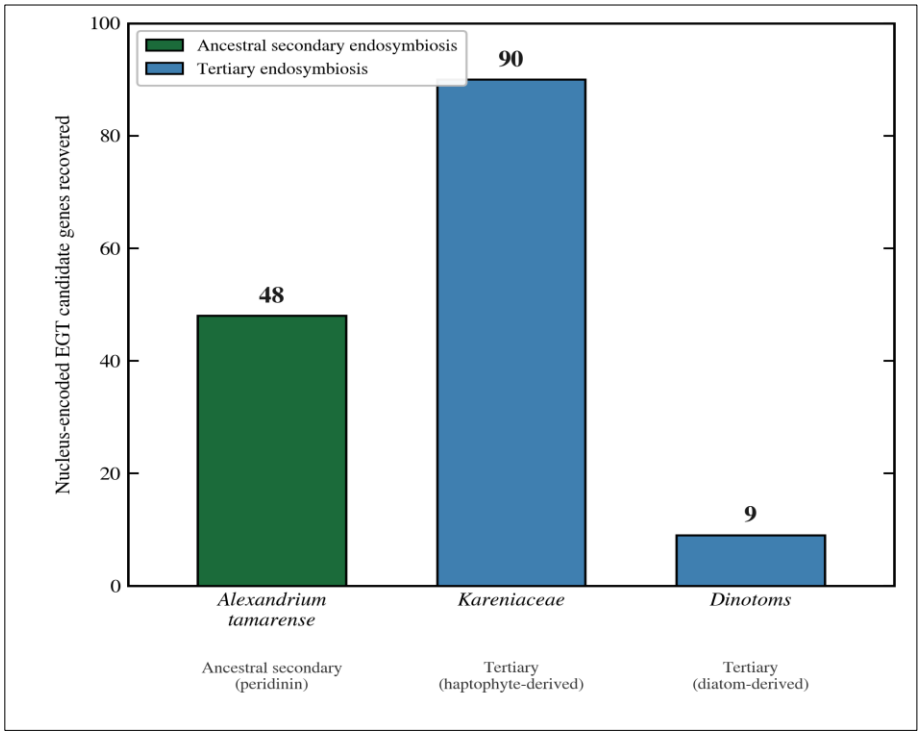

Fig 4 Magnitude of Red-/Donor-Derived Gene Transfer to the Dinoflagellate Nucleus

Ancestral secondary (peridinin) transfer in *Alexandrium tamarense* versus tertiary transfer in haptophyte-derived Kareniaceae and diatom-derived dinotoms [7, 12]. The counts are related but not identical, as noted in the text.

➢ *Methodological Heterogeneity Among Studies*

The included studies vary considerably in rigour, and this variation is the main source of apparent disagreement in the literature. Early reports relied on similarity (best-hit) searches with BLAST that could not distinguish EGT from HGT or contamination, and were never graded above weak [28]. The strong grades were reserved for studies that constructed explicit single- or multi-gene phylogenies with adequate sampling of red algae and competitors [4, 25, 33]. Targeting prediction also varied: some studies validated localisation experimentally, whereas others relied on predictors of limited accuracy in dinoflagellates, owing to the weakly conserved bipartite targeting motif [18, 20, 21]. Reference databases ranged from pre-genomic sampling to the MMETSP and EukProt collections; the progression from EST to transcriptome to genome data improved completeness but also increased the risk of contamination from symbionts and prey, a risk explicitly mitigated only in more recent genome-scale studies [12, 14, 16]. Because of these differences, apparent agreement between studies can be misleading, as they may in fact differ in methodology rather than in conclusion. The per-study appraisal across the seven domains of Section 2.9 is summarised in Figure 2.

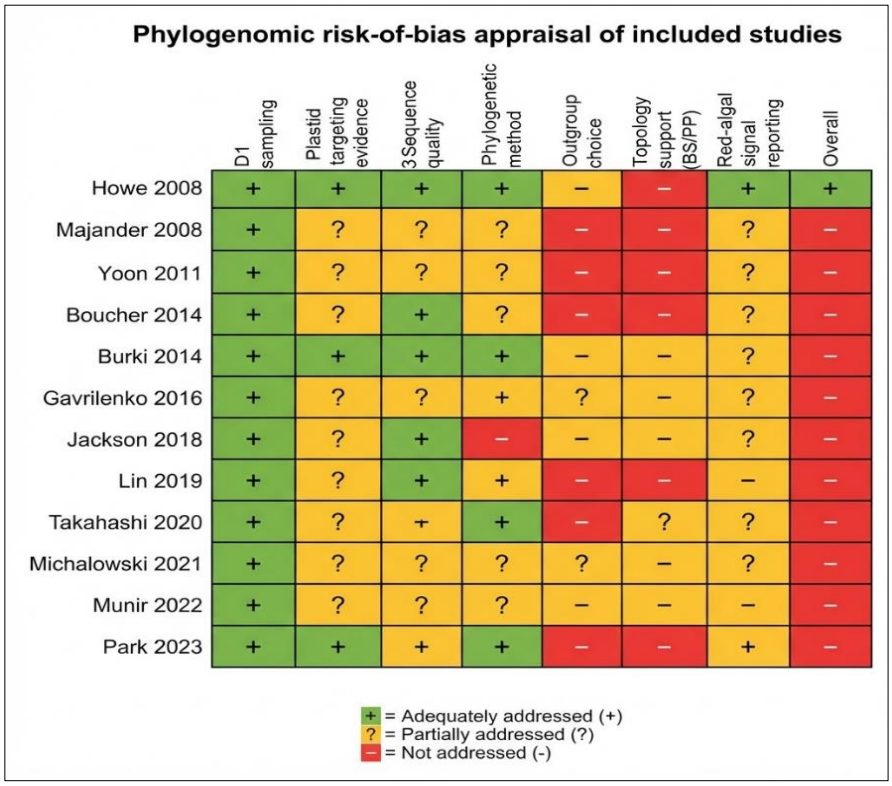

Fig 5 Phylogenomic Risk-of-Bias Appraisal of the Included Studies (Traffic-Light Plot).

Seven adapted appraisal domains (D1–D7) and an overall judgement are scored for each study as adequately addressed (+), partially addressed (?) or not addressed (−). This evolutionary-genomics appraisal is the legitimate analogue of a clinical risk-of-bias assessment; the Cochrane RoB2 tool and a forest plot were not applicable because no randomised trials and no poolable quantitative outcomes were included.

IV. DISCUSSION

➢ *Summary of Main Findings*

This systematic review shows that the evidence for red-algal EGT in dinoflagellates is real and repeatedly recovered, but strongly constrained in both function and evolutionary history. Nucleus-encoded proteins of the photosynthetic core, the Calvin–Benson cycle and plastid gene expression are supported with high confidence, whereas peripheral metabolic, transport and pigment functions are not. Confidence is likewise highest in peridinin lineages, where the red signal is inherited directly from the ancestral secondary endosymbiosis, and lower in lineages that have replaced their plastid, where younger genes partly overwrite the original signal [9, 12]. This asymmetry is quantified by the order-of-magnitude difference between the 48 ancestrally derived nuclear genes recovered in *Alexandrium* and the 9 to 90 donor-derived genes recovered in dinotoms and Kareniaceae, respectively [7, 12]. In short, red-algal EGT is a major, functionally biased contributor to

the ancestral secondary plastid and a minor, donor-specific contributor to the later tertiary plastids.

➢ *Evolutionary Implications*

The pattern recovered here agrees with the now-broad consensus that the plastids of apicomplexans, heterokonts and dinoflagellates originated from the same red-algal secondary endosymbiosis [1, 4, 41]. Under this scenario, the dinoflagellate ancestor (or its host) acquired a large complement of plastid-targeted genes early in dinoflagellate evolution, and these genes persist despite the later fragmentation of the plastid genome into minicircles [6, 7]. The mosaic ancestry of many proteins is therefore not evidence against a red origin, but evidence of continued layering of additional acquisitions — from green algae, bacteria and tertiary donors — upon an essentially red foundation [10, 13, 22]. Dinoflagellates are among the clearest examples of plastid retention, replacement and loss in eukaryote evolution [1], and the relic-nucleus dinotoms uniquely preserve a living snapshot of the early stages of plastid integration and ongoing EGT [37].

A further implication concerns how EGT scales from secondary to tertiary endosymbiosis. The much larger number of genes transferred to the host nucleus during the ancestral event implies that most functionally essential plastid genes had already been transferred and stabilised in the dinoflagellate nucleus before the tertiary endosymbioses occurred. A replacement plastid could therefore be maintained and supplied largely by genes already present in the host, requiring comparatively few new transfers [11, 12]. This "pre-loaded nucleus" may help to explain why plastid replacement is so common in dinoflagellates: because the host nucleus already encodes a large fraction of the plastid proteome, a new organelle can be acquired and integrated more readily [5].

➢ *Functional Implications*

The functional bias documented in Tables 4 and 5 is biologically meaningful. The proteins with the strongest red-algal signal are those essential for carbon fixation, the light reactions and the constant co-regulation of plastid function — GAPDH, FBPase, SBPase, photosystem subunits and oxygen-evolving enhancer proteins — which were likely transferred to the nucleus early and permanently [25, 33]. Genes of this kind allow the host to control photosynthate production, pigment synthesis and redox balance, and render it dependent on the plastid for primary metabolism [5, 16]. In the symbiotic Symbiodiniaceae this integration extends to the animal host, and genome analyses have identified expansions and positive selection among genes involved in the coral–algal partnership [17]. The weaker, more mixed signal in peripheral pathways indicates that not all plastid functions require a red-algal gene, reinforcing the view that the dinoflagellate plastid proteome is a functional composite rather than a simple collection of red-algal genes [10].

➢ *Methodological Challenges*

Several recurrent problems affect the reliability of EGT inference in dinoflagellates, and each must be addressed in a credible study. First, genome complexity: because dinoflagellate genomes are frequently represented only by incomplete assemblies or by transcriptomes, apparent loss of a gene from the organellar genome cannot be established with confidence [15, 16]. Second, contamination: many datasets derive from non-axenic cultures or from symbiotic associations, so sequences from prey, bacteria or animal hosts can masquerade as host sequences, a problem that is especially acute for transcriptome-only analyses [12, 14]. Third, phylogenetic artefact: the rapidly evolving plastid-derived genes of dinoflagellates are prone to long-branch attraction, which can group divergent dinoflagellate sequences with unrelated long-branch taxa and misassign the donor [5]. Fourth, hidden paralogy arising from ancient duplications can produce orthologue misidentification and misleading topologies. Fifth, reference sparsity: red-algal genomes are poorly represented relative to green algae and bacteria, so a genuine red affinity may go undetected when the corresponding rhodophyte sequences are absent from the database [4]. Sixth, prediction: the weakly conserved bipartite targeting motif of peridinin dinoflagellates reduces predictive accuracy, so a protein may be incorrectly classified as plastid-targeted or not [18, 21]. Finally, the entanglement of EGT with HGT — illustrated by bacterial form II Rubisco and by green PRK — means that a red-algal EGT claim is secure only when alternative origins for these enzymes have been tested and rejected [4, 13].

➢ *Recommendations for Future EGT Inference*

Given the heterogeneity described above, and on the principle that any claim of red-algal EGT in a dinoflagellate should meet at least the minimum requirements set out in Table 2, we recommend that these requirements be reported explicitly for every future claim. Studies should search widely for homologues across red algae, green algae, other red-lineage groups, bacteria and eukaryotic outgroups using sensitive tools [28, 29]; assess orthology and paralogy explicitly; predict plastid targeting with a red-lineage-aware tool and, where possible, validate the prediction experimentally [18, 32]; align and trim reproducibly [30, 31]; infer trees under model selection with both maximum-likelihood and, where feasible, Bayesian support; and make alignments and trees publicly available for independent audit. Many of the weak and ambiguous grades in Table 4 would become defensible assignments if such a framework were combined with more complete reference datasets for red algae and alveolates.

➢ *Limitations and Future Research Directions*

Progress will depend on closing the gaps identified above. Comparative plastid proteomics that combines prediction with mass-spectrometric verification would replace inferred targeting with direct evidence. Beyond the Symbiodiniaceae, EGT could be quantified using additional high-quality, chromosome-scale dinoflagellate genomes

spanning peridinin, kareniacean, dinotom and heterotrophic lineages [16, 17, 37]; non-photosynthetic and heterotrophic dinoflagellates, at present almost entirely unsampled, are a particular priority [39]. Single-cell genomics offers a route to the uncultured majority, and higher-quality genomes for red algae and alveolates would directly improve donor discrimination. Targeting, and ideally the function of the gene product, should be tested experimentally wherever possible. Above all, the field would benefit from phylogenomic pipelines designed specifically for EGT inference, incorporating contamination screening, long-branch mitigation, explicit orthology resolution and traceable evidence grading [5, 12]. Like the primary literature it summarises, the present review is itself constrained by uneven taxon sampling and by the limited number of high-quality reference genomes currently available, and its conclusions should be revisited as these resources expand.

V. CONCLUSIONS

Red-algal endosymbiotic gene transfer is a genuine, extensive and functionally central component of the nuclear genome of secondary plastid-bearing dinoflagellates. The strongest and most reproducible evidence concerns nucleus-encoded proteins of the light reactions, the Calvin–Benson cycle and plastid gene expression, and is clearest in peridinin lineages that still carry the ancestral red plastid inherited from the chromalveolate secondary endosymbiosis [4, 7, 33]. In the Kareniaceae and dinotoms, plastid replacement superimposes a younger, donor-specific and comparatively shallow layer of transfer that often obscures the ancestral signal, giving rise to the mosaic proteomes reported for many dinoflagellate plastid proteins [9, 10, 12]. Where confidence is currently low, the cause is more often methodological than biological — few reference sequences for red algae and alveolates, transcriptome contamination, long-branch and paralogy artefacts, imperfect targeting prediction, and the continuing entanglement of EGT with HGT. Standardised and transparent inference frameworks, more complete reference genomes, experimental validation and EGT-specific phylogenomic pipelines are therefore the key priorities for future work. With these tools, the dinoflagellate nucleus can be read not merely as a repository of disarticulated red-algal genes, but as a remarkably well-preserved and legible record of the dismantling, relocation and repeated replacement of a red-algal endosymbiont.

REFERENCES

[1]. Keeling, P. J. (2010). The endosymbiotic origin, diversification and fate of plastids. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1541), 729–748. https://doi.org/10.1098/rstb.2009.0103

[2]. Archibald, J. M. (2015). Genomic perspectives on the birth and spread of plastids. *Proceedings of the National Academy of Sciences*, *112*(33), 10147–10153. https://doi.org/10.1073/pnas.1421374112

[3]. Dorrell, R. G., & Howe, C. J. (2012). What makes a chloroplast? Reconstructing the establishment of photosynthetic symbioses. *Journal of Cell Science*, *125*(8), 1865–1875. https://doi.org/10.1242/jcs.102285

[4]. Janouškovec, J., Horák, A., Oborník, M., Lukeš, J., & Keeling, P. J. (2010). A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. *Proceedings of the National Academy of Sciences*, *107*(24), 10949–10954. https://doi.org/10.1073/pnas.1003335107

[5]. Dorrell, R. G., & Howe, C. J. (2015). Integration of plastids with their hosts: Lessons learned from dinoflagellates. *Proceedings of the National Academy of Sciences*, *112*(33), 10247–10254. https://doi.org/10.1073/pnas.1421380112

[6]. Howe, C. J., Nisbet, R. E. R., & Barbrook, A. C. (2008). The remarkable chloroplast genome of dinoflagellates. *Journal of Experimental Botany*, *59*(5), 1035–1045. https://doi.org/10.1093/jxb/erm292

[7]. Hackett, J. D., Yoon, H. S., Soares, M. B., Bonaldo, M. F., Casavant, T. L., Scheetz, T. E., Nosenko, T., & Bhattacharya, D. (2004). Migration of the plastid genome to the nucleus in a peridinin dinoflagellate. *Current Biology*, *14*(3), 213–218. https://doi.org/10.1016/j.cub.2004.01.032

[8]. Gabrielsen, T. M., Minge, M. A., Espelund, M., Tooming-Klunderud, A., Patil, V., Nederbragt, A. J., Otis, C., Turmel, M., Shalchian-Tabrizi, K., Lemieux, C., & Jakobsen, K. S. (2011). Genome evolution of a tertiary dinoflagellate plastid. *PLoS ONE*, *6*(4), e19132. https://doi.org/10.1371/journal.pone.0019132

[9]. Ishida, K., & Green, B. R. (2002). Second- and third-hand chloroplasts in dinoflagellates: Phylogeny of oxygen-evolving enhancer 1 (PsbO) protein reveals replacement of a nuclear-encoded plastid gene by that of a haptophyte tertiary endosymbiont. *Proceedings of the National Academy of Sciences*, *99*(14), 9294–9299. https://doi.org/10.1073/pnas.142091799

[10]. Minge, M. A., Shalchian-Tabrizi, K., Tørresen, O. K., Takishita, K., Probert, I., Inagaki, Y., Klaveness, D., & Jakobsen, K. S. (2010). A phylogenetic mosaic plastid proteome and unusual plastid-targeting signals in the green-colored dinoflagellate *Lepidodinium chlorophorum*. *BMC Evolutionary Biology*, *10*, 191. https://doi.org/10.1186/1471-2148-10-191

[11]. Patron, N. J., Waller, R. F., & Keeling, P. J. (2006). A tertiary plastid uses genes from two endosymbionts. *Journal of Molecular Biology*, *357*(5), 1373–1382. https://doi.org/10.1016/j.jmb.2006.01.084

[12]. Burki, F., Imanian, B., Hehenberger, E., Hirakawa, Y., Maruyama, S., & Keeling, P. J. (2014). Endosymbiotic gene transfer in tertiary plastid-

containing dinoflagellates. *Eukaryotic Cell*, *13*(2), 246–255. https://doi.org/10.1128/EC.00299-13

[13]. Nosenko, T., Lidie, K. L., Van Dolah, F. M., Lindquist, E., Cheng, J.-F., & Bhattacharya, D. (2006). Chimeric plastid proteome in the Florida "red tide" dinoflagellate *Karenia brevis*. *Molecular Biology and Evolution*, *23*(11), 2026–2038. https://doi.org/10.1093/molbev/msl074

[14]. Keeling, P. J., Burki, F., Wilcox, H. M., Allam, B., Allen, E. E., Amaral-Zettler, L. A., … Worden, A. Z. (2014). The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biology*, *12*(6), e1001889. https://doi.org/10.1371/journal.pbio.1001889

[15]. Shoguchi, E., Shinzato, C., Kawashima, T., Gyoja, F., Mungpakdee, S., Koyanagi, R., … Satoh, N. (2013). Draft assembly of the *Symbiodinium minutum* nuclear genome reveals dinoflagellate gene structure. *Current Biology*, *23*(15), 1399–1408. https://doi.org/10.1016/j.cub.2013.05.062

[16]. Lin, S., Cheng, S., Song, B., Zhong, X., Lin, X., Li, W., … Morse, D. (2015). The *Symbiodinium kawagutii* genome illuminates dinoflagellate gene expression and coral symbiosis. *Science*, *350*(6261), 691–694. https://doi.org/10.1126/science.aad0408

[17]. Liu, H., Stephens, T. G., González-Pech, R. A., Beltran, V. H., Lapeyre, B., Bongaerts, P., … Chan, C. X. (2018). *Symbiodinium* genomes reveal adaptive evolution of functions related to coral–dinoflagellate symbiosis. *Communications Biology*, *1*, 95. https://doi.org/10.1038/s42003-018-0098-3

[18]. Gruber, A., Rocap, G., Kroth, P. G., Armbrust, E. V., & Mock, T. (2015). Plastid proteome prediction for diatoms and other algae with secondary plastids of the red lineage. *The Plant Journal*, *81*(3), 519–528. https://doi.org/10.1111/tpj.12734

[19]. Nassoury, N., Cappadocia, M., & Morse, D. (2003). Plastid ultrastructure defines the protein import pathway in dinoflagellates. *Journal of Cell Science*, *116*(14), 2867–2874. https://doi.org/10.1242/jcs.00517

[20]. Patron, N. J., Waller, R. F., Archibald, J. M., & Keeling, P. J. (2005). Complex protein targeting to dinoflagellate plastids. *Journal of Molecular Biology*, *348*(4), 1015–1024. https://doi.org/10.1016/j.jmb.2005.03.030

[21]. Patron, N. J., & Waller, R. F. (2007). Transit peptide diversity and divergence: A global analysis of plastid targeting signals. *BioEssays*, *29*(10), 1048–1058. https://doi.org/10.1002/bies.20638

[22]. Dorrell, R. G., Gile, G., McCallum, G., Méheust, R., Bapteste, E. P., Klinger, C. M., Brillet-Guéguen, L., Freeman, K. D., Richter, D. J., & Bowler, C. (2017). Chimeric origins of ochrophytes and haptophytes revealed through an ancient plastid proteome. *eLife*, *6*, e23717. https://doi.org/10.7554/eLife.23717

[23]. Harper, J. T., Waanders, E., & Keeling, P. J. (2005). On the monophyly of chromalveolates using a six-protein phylogeny of eukaryotes. *International Journal of Systematic and Evolutionary Microbiology*, *55*(1), 487–496. https://doi.org/10.1099/ijs.0.63216-0

[24]. Fast, N. M., Kissinger, J. C., Roos, D. S., & Keeling, P. J. (2001). Nuclear-encoded, plastid-targeted genes suggest a single common origin for apicomplexan and dinoflagellate plastids. *Molecular Biology and Evolution*, *18*(3), 418–426. https://doi.org/10.1093/oxfordjournals.molbev.a003818

[25]. Harper, J. T., & Keeling, P. J. (2003). Nucleus-encoded, plastid-targeted glyceraldehyde-3-phosphate dehydrogenase (GAPDH) indicates a single origin for chromalveolate plastids. *Molecular Biology and Evolution*, *20*(10), 1730–1735. https://doi.org/10.1093/molbev/msg195

[26]. Wang, Y., Joly, S., & Morse, D. (2008). Phylogeny of dinoflagellate plastid genes recently transferred to the nucleus supports a common ancestry with red algal plastid genes. *Journal of Molecular Evolution*, *66*(2), 175–184. https://doi.org/10.1007/s00239-008-9070-z

[27]. Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, n71. https://doi.org/10.1136/bmj.n71

[28]. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

[29]. Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, *12*(1), 59–60. https://doi.org/10.1038/nmeth.3176

[30]. Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, *25*(15), 1972–1973. https://doi.org/10.1093/bioinformatics/btp348

[31]. Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

[32]. Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., von Heijne, G., & Nielsen, H. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology*, *37*(4), 420–423. https://doi.org/10.1038/s41587-019-0036-z

[33]. Teich, R., Zauner, S., Baurain, D., Brinkmann, H., & Petersen, J. (2007). Origin and distribution of Calvin cycle fructose and sedoheptulose bisphosphatases in Plantae and complex algae: A single secondary origin of complex red plastids and subsequent propagation

via tertiary endosymbioses. *Protist, 158*(3), 263–276. https://doi.org/10.1016/j.protis.2006.12.004

[34]. Richardson, E., Dorrell, R. G., & Howe, C. J. (2014). Genome-wide transcript profiling reveals the coevolution of plastid gene sequences and transcript processing pathways in the fucoxanthin dinoflagellate *Karlodinium veneficum*. *Molecular Biology and Evolution, 31*(9), 2376–2386. https://doi.org/10.1093/molbev/msu189

[35]. González-Pech, R. A., Ragan, M. A., & Chan, C. X. (2017). Signatures of adaptation and symbiosis in genomes and transcriptomes of *Symbiodinium*. *Scientific Reports, 7*, 15021. https://doi.org/10.1038/s41598-017-15029-w

[36]. LaJeunesse, T. C., Parkinson, J. E., Gabrielson, P. W., Jeong, H. J., Reimer, J. D., Voolstra, C. R., & Santos, S. R. (2018). Systematic revision of Symbiodiniaceae highlights the antiquity and diversity of coral endosymbionts. *Current Biology, 28*(16), 2570–2580.e6. https://doi.org/10.1016/j.cub.2018.07.008

[37]. Sarai, C., Tanifuji, G., Nakayama, T., Kamikawa, R., Takahashi, K., Yazaki, E., … Tanaka, S. (2020). Dinoflagellates with relic endosymbiont nuclei as models for elucidating organellogenesis. *Proceedings of the National Academy of Sciences, 117*(10), 5364–5375. https://doi.org/10.1073/pnas.1911884117

[38]. Lo, R., Dougan, K. E., Chen, Y., Shah, S., Bhattacharya, D., & Chan, C. X. (2022). Alignment-free analysis of whole-genome sequences from Symbiodiniaceae reveals different phylogenetic signals in distinct regions. *Frontiers in Plant Science, 13*, 815714. https://doi.org/10.3389/fpls.2022.815714

[39]. Cooney, E. C., Holt, C. C., Hehenberger, E., Adams, J. A., Leander, B. S., & Keeling, P. J. (2024). Investigation of heterotrophs reveals new insights in dinoflagellate evolution. *Molecular Phylogenetics and Evolution, 196*, 108086. https://doi.org/10.1016/j.ympev.2024.108086

[40]. Ponce-Toledo, R. I., Moreira, D., López-García, P., & Deschamps, P. (2018). Secondary plastids of euglenids and chlorarachniophytes function with a mix of genes of red and green algal ancestry. *Molecular Biology and Evolution, 35*(9), 2198–2204. https://doi.org/10.1093/molbev/msy121

[41]. Burki, F., Roger, A. J., Brown, M. W., & Simpson, A. G. B. (2020). The new tree of eukaryotes. *Trends in Ecology & Evolution, 35*(1), 43–55. https://doi.org/10.1016/j.tree.2019.08.008