



The Integration of Multiple Nuclear-Encoded Transgenes in the Green Alga *Chlamydomonas reinhardtii* Results in Higher Transcription Levels

Noam Shahar¹, Shira Landman¹, Iddo Weiner^{1,2}, Tamar Elman¹, Eyal Dafni¹, Yael Feldman¹, Tamir Tuller^{2,3} and Iftach Yacoby^{1*}

OPEN ACCESS

Edited by:

Dimitris Petroustos,
UMR5168 Laboratoire de Physiologie
Cellulaire Vegetale (LPCV), France

Reviewed by:

Ales Kovarik,
Academy of Sciences of the Czech
Republic (ASCR), Czechia
Pierre Crozet,
Sorbonne Universités, France
Lutz Wobbe,
Bielefeld University, Germany

*Correspondence:

Iftach Yacoby
iftachy@tauex.tau.ac.il

Specialty section:

This article was submitted to
Plant Biotechnology,
a section of the journal
Frontiers in Plant Science

Received: 26 August 2019

Accepted: 20 December 2019

Published: 14 February 2020

Citation:

Shahar N, Landman S, Weiner I,
Elman T, Dafni E, Feldman Y, Tuller T
and Yacoby I (2020) The Integration of
Multiple Nuclear-Encoded Transgenes
in the Green Alga *Chlamydomonas*
reinhardtii Results in Higher
Transcription Levels.
Front. Plant Sci. 10:1784.
doi: 10.3389/fpls.2019.01784

¹ The George S. Wise Faculty of Life Sciences, School of Plant Sciences and Food Security, Tel Aviv University, Tel Aviv, Israel, ² Department of Biomedical Engineering, The Iby and Aladar Fleischman Faculty of Engineering, Tel Aviv University, Tel Aviv, Israel, ³ The Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel

The integration of genes into the nuclear genome of *Chlamydomonas reinhardtii* is mediated by Non-Homologous-End-Joining, thus resulting in unpredicted insertion locations. This phenomenon defines ‘the position-effect’, which is used to explain the variation of expression levels between different clones transformed with the same DNA fragment. Likewise, nuclear transgenes often undergo epigenetic silencing that reduces their expression; hence, nuclear transformations require high-throughput screening methods to isolate clones that express the foreign gene at a desirable level. Here, we show that the number of integration sites of heterologous genes results in higher mRNA levels. By transforming both a synthetic ferredoxin-hydrogenase fusion enzyme and a Gaussia-Luciferase reporter protein, we were able to obtain 33 positive clones that exhibit a wide range of synthetic expression. We then performed a droplet-digital polymerase-chain-reaction for these lines to measure their transgene DNA copy-number and mRNA levels. Surprisingly, most clones contain two integration sites of the synthetic gene (45.5%), whilst 33.3% contain one, 18.1% include three and 3.1% encompass four. Remarkably, we observed a positive correlation between the raw DNA copy-number values to the mRNA levels, suggesting a general effect of which transcription of transgenes is partially modulated by their number of copies in the genome. However, our data indicate that only clones harboring at least three copies of the target amplicon show a significant increment in mRNA levels of the reporter transgene. Lastly, we measured protein activity for each of the reporter genes to elucidate the effect of copy-number variation on heterologous expression.

Keywords: copy-number variation, *Chlamydomonas reinhardtii*, Microalgae, the position-effect, droplet-digital PCR

INTRODUCTION

Two main approaches are routinely used for transforming the green alga *Chlamydomonas reinhardtii*; the first includes integration into the chloroplast genome whilst the second includes integration into the nucleus. As opposed to its plastomic counterpart, nuclear transformation of *C. reinhardtii* is practically easier to perform and does not require selective cycles to obtain transgene homoplasmy (Yamano et al., 2013). However, the integration of synthetic genes into the nuclear genome predominantly occurs randomly *via* Non-Homologous-End-Joining (NHEJ), leading to a large population of transformed cells with varied expression levels (Yamano et al., 2013; Weiner et al., 2018). Such phenomenon requires high-throughput screening methods for isolation of clones expressing the recombinant protein of interest at a desirable level, a non-trivial task that varies between target transgenes.

Such heterogeneity of nucleus-transformed clones is mainly explained by the “position effect”, a common phenomenon in which the expression of foreign genes is modulated by the genomic properties of their integration sites (León and Fernández, 2007; Zhang et al., 2014). It was previously shown that in order to average out position effects, one should collect at least 240 transformed colonies of *C. reinhardtii* (Lodha et al., 2008). Moreover, it was shown that the integration of plasmids occasionally results in deletions of genomic sequences, genomic rearrangements, and insertion of short DNA sequences (derived from fragmented DNA of lysed cells that are co-transformed along with the cassette) (Zhang et al., 2014; Jinkerson and Jonikas, 2015). Furthermore, gene silencing is an additional factor that plays a role in hindering heterologous expression in microalgae; it was shown to occur both at the transcriptional and post-transcriptional levels (Wu-scharf and Cerutti, 2000; Shaver et al., 2010), and is most likely part of an “immune system” protecting the cell against viral infections and transposable elements. Overall, the process of obtaining nuclear-transformed clones with a reasonably high expression of synthetic genes in *C. reinhardtii* is challenging.

In attempts to overcome these limitations, recent studies have developed different genetic approaches; for example, by performing an Ultraviolet (UV) mutagenesis on transformed clones with low transgene expression levels, two clones that further displayed a significant increase in the accumulation of the foreign gene were isolated (UVM-4 and UVM-11) (Neupert et al., 2009). Additional transformations into these clones resulted in a relatively high and uniform expression of heterologous genes (Barahimipour et al., 2015; Barahimipour et al., 2016). It was demonstrated that this unique phenotype of UVM-4 and UVM-11 is due to a gain-of-function mutation in the chromatin modification gene H4Ac, which most likely plays a key-role in repressing the chromatin state at the time of exogenous DNA integration (Barahimipour et al., 2016). Another approach used is the development of a detection system for measuring the frequency of homologous recombination (HR) events to successfully select clones with a desired integration site (Nour-Eldin et al., 2018). In yet another path, sequence optimization techniques were applied on target

genes to mimic endogenous features of ORFs (e.g., the particularly high GC-content of the *Chlamydomonas* nuclear genome and the introduction of native introns into transgenes), and by that reducing the number of false clones (i.e., resistant to the selective pressure, but do not express the target transgene) and increase overall heterologous expression (Baier et al., 2018; Weiner et al., 2018). Furthermore, the development of the CRISPR/Cas9 genetic system in *C. reinhardtii* could potentially reduce the position-effect by “directing” the inserted transgene into a specific locus (Shin et al., 2016; Greiner et al., 2017).

Another important genetic feature that is known to create population diversity is the Copy-Number Variation (CNV); it is defined as a DNA segment found at variable copy quantities in a homogenic population of cells. CNV is linked to dozens of diseases (Redon et al., 2006; Malhotra and Sebat, 2012), but also serves as an important means to create and conserve population heterogeneity (Redon et al., 2006). In the context of synthetic biology, and especially in plants — where transgenes are mostly integrated randomly into the nuclear genome — it was shown that heterologous genes can be inserted more than once into different genomic locations, or as tandem repeats at a single locus to play either a positive or negative role on their expression (Stockhaus et al., 1987; Hobbs et al., 1990; De Block, 1993; Hobbs et al., 1993; Jorgensen et al., 1996; Que et al., 1997; Ku et al., 1999; Schubert et al., 2004; Wang et al., 2015; Zhang et al., 2015); while integration into distant loci mainly results in a synergistic effect (i.e., the expression is increased proportionally to its cognate DNA copy-number), repeat arrangements often induce transgene-silencing. However, a comprehensive study by Schubert et al., which was carried on 132 transgenic lines of *Arabidopsis thaliana* using three different reporters, emphasized the positive effect of copy-number on expression levels regardless to the genomic position and/or arrangement (repeats or inverted repeats) (Schubert et al., 2004). Generally, copy-number analysis in transgenic plants is carried out to select for lines that harbor a single copy of the integrated gene, as these typically segregate in a Mendelian fashion and are less prone to exhibit gene-silencing (Wang et al., 2015). Hence, such synthetic-CNV (sCNV) can be a major element that increases variation of nuclear transgenic lines in *C. reinhardtii* as well. For example, it was previously suggested that strategies that intentionally tend to increase transgenes copy-number may improve synthetic expression in *C. reinhardtii* (Lauersen et al., 2016). Another study indirectly approved the positive effect of transgenes copy-number by transforming *C. reinhardtii* with several different plasmids that contained the same synthetic gene, fused onto different fluorescence proteins (Wichmann et al., 2018). However, a large-scale analysis of sCNV has not yet been studied in *C. reinhardtii* or in any other microalga.

Today, several methods are used for copy-number analysis: (i) the traditional southern-blot, (ii) Quantitative PCR (qPCR), and (iii) droplet-digital PCR (ddPCR) (Southern, 1975; Schloss, 1990; Hindson et al., 2011; Wang et al., 2015; Zhang et al., 2015; Stefano et al., 2016; Collier et al., 2017). While southern-blot is a powerful tool (e.g., it provides additional information regarding the genomic organization and the integrity of the integrated

cassette), it is arduous to use for a large number of samples, as it involves many laborious and time consuming steps, and was shown to yield unpredictable results when high number of DNA copies are examined (Collier et al., 2017). The recently developed ddPCR method was shown to detect the copy-number of genes as accurately as southern-blot analysis (Hindson et al., 2011; Pinheiro et al., 2012; Collier et al., 2017), and can be performed on a multiple number of samples simultaneously. Moreover, ddPCR has been optimized for CNV analysis in crop-plants (Demeke et al., 2014; Collier et al., 2017); for these reasons, we used ddPCR analysis to examine sCNV in *C. reinhardtii*.

In this work we studied the correlation between the number of DNA integration events, mRNA levels and expression levels in engineered strains of *C. reinhardtii*. For that purpose, we transformed both a synthetic ferredoxin-hydrogenase fusion enzyme (fd-hyd) (Eilenberg et al., 2016; Weiner et al., 2018) and a Gaussia-Luciferase reporter protein (gLuc) (Shao and Bock, 2008) into the nuclear genome of *C. reinhardtii*. We obtained 33 clones that exhibit a wide range of synthetic expression levels, and subsequently performed each ddPCR to measure transgene DNA copy-number and mRNA levels. Finally, we measured protein activity for each of these clones to elucidate the impact of sCNV on heterologous expression.

MATERIALS AND METHODS

Algal Cultures, Growth Conditions and Chlorophyll Determination

The algal cultures of the wild-type clone CC-124 and the hyda_{1,2} double-mutant (Meuser et al., 2012) were grown in Tris-Acetate-Phosphate (TAP) medium at 25°C under continuous cool daylight and cool white fluorescent lights (90 μmol m⁻² s⁻¹) stirring in 100 ml Erlenmeyers capped with silicone sponge enclosures. Experiments were performed when the cultures reached the density of 3 × 10⁶ cells/ml (as determined by Celeromics cell counter) that corresponds to chlorophyll concentrations of ~9 μg Chl/ml.

C. reinhardtii Nuclear Transformation

DNA sequences were synthesized and cloned (using a recombination-based technique - CloneEZ™), into the pSL18::aph7 (Fischer and Rochaix, 2001) or pChlamy_1 (GeneArt™) plasmids, under the control of the endogenous *PSAD* or *HSP70A/RBCS2* promoters for fd-hyd and gLuc, respectively. 2 μg of linearized fd-hyd and gLuc plasmids were transformed by electroporation into 1.5 × 10⁷ cells of hyda_{1,2} double-mutant knockout strain and the wild-type strain CC-124, respectively, following the GeneArt *Chlamydomonas* Engineering Kit protocol (Life Technologies). Initial screening was carried out on Hygromycin (10 μg/ml).

Rhodobacter capsulatus High Throughput Screen

Prepared precisely as described in (Wecker et al., 2011). Algal strains, overlaid with engineered H₂-sensing *R. capsulatus*, were

scanned using the Fuji FLA-5100 fluorescence imager. A 473 nm laser was used for excitation whereas 510 nm or 665 nm filters were used for quantifying GFP fluorescence and chlorophyll density, respectively.

Luciferase *In Vivo* Assay

The protocol for Luciferase *in vivo* screening was adopted from (Mayfield and Schultz, 2004; Shao et al., 2007; Shao and Bock, 2008; Rasala et al., 2011) with the following modifications: the algal colonies were incubated for 10 minutes in dark, and then 10 μl of 0.01 mM Coelenterazine were dropped onto each colony. Finally, the plate was scanned for bioluminescence using the MicroChemi 6.0 camera.

Nucleic Acids Extraction

One hundred milligrams of the cell pellet were taken for total RNA extraction using a RNeasy Plant Mini Kit (Qiagen 74903, <http://www.qiagen.com/>). One microgram of purified RNA from each sample was used for synthesis of complementary DNA (cDNA) using Applied Bio-System High-Capacity cDNA Reverse-Transcription Kit (with random primers). Thirty milligrams of the cell pellet were taken for total DNA extraction using an OMEGA E.Z.N.A SP Plant DNA kit. DNA and RNA concentrations were quantified by a Thermo Scientific™ NanoDrop™ 2000 Spectrophotometer.

Droplet Digital PCR

The ddPCR reaction mixture was prepared from a QX200™ ddPCR™ EvaGreen Supermix, 250 nM primers, 5 ng of cDNA/1 ng of DNA template (the cDNA quantity was based on the initial amount of RNA concentration used for the reverse-complementation) and DDW to a total volume of 22 μl. Each reaction mixture was loaded into a DG8™ Cartridge (Bio-Rad), and a volume of 70 μl of QX200™ Droplet Generation Oil was loaded into the oil-well of each channel. The DG8™ Cartridge was then placed into the QX200™ Droplet Generator. Next, 40 μl of the created droplets were transferred to a ddPCR™ 96-Well Plate. The plate was heat-sealed with a foil seal and further undergone PCR using C1000 Touch™ Thermal Cycler. Subsequently, the 96-well PCR plate was placed onto the QX200™ Droplet Reader. Further analysis of the data was performed using QuantaSoft™, where the absolute concentration of each target (fd-hyd or gLuc) was normalized to the absolute quantification of the reference gene *CBLP*. The primers used were:

- fd-hyd (5'- TCGCCATGCTGGAGAAGTC -3', 5'- AGCTGGACACGTAGGGGAT -3');
- gLuc (5'- CCCAAGTGGACCTGTGTGTG -3', 5'- CACTTCTTCAGCAGGTCGCTA -3');
- *CBLP* (5'- AAAGCCACGGACAGCACATC-3', 5'- GATGGCCAGTTCTGCTGAC-3').

fd-hyd Induction

Anaerobic induction was carried out as follows: liquid cell cultures (40 ml) at mid log phase (3 × 10⁶ cells/ml) were

concentrated by centrifugation (2,800 g for 7 min) and re-suspended in 4 ml of TAP HEPES buffer (50 mM HEPES, pH 7.8). The concentrated cells were transferred into septum stopper sealed 14 ml glass serum vials (Wheaton), covered by aluminum foil and purged for 30 min with Argon at room temperature. Thereafter, the concentrated cultures were incubated in dark, at room temperature, under 60 rpm agitation for additional 90 minutes.

Hydrogenase Activity Assay by Methyl-Viologen

Carried out precisely as described in (Eilenberg et al., 2016). Following two hours of dark anaerobiosis, cells were transferred into a buffer containing reduced methyl viologen and Triton-X for lysing the cells. A 500 μ l sample was drawn from the headspace and the H₂ concentration was determined by gas chromatography. The amount of enzyme was calculated based on the constant fd-hyd specific activity (Yacoby et al., 2011).

Statistical Analysis

Pearson Correlation

Pearson correlation coefficients (ρ) and p-values were calculated using python's "scipy.stats.pearsonr" function.

Two-Way ANOVA

To determine whether the effect of both the reporter type (fd-hyd or gLuc) and the DNA copy-number on the mRNA levels is significant, a two-way ANOVA test was carried out using SPSS.

Tukey's Range Test

To determine which of the copy-number group mRNA levels means are statistically different from each other, a Tukey's range test was performed using SPSS.

RESULTS

All data are summarized in **Table S1**.

Construction and Isolation of an Array of Clones With Varied Expression Levels

To obtain clones with a wide-range of synthetic expression levels, we took advantage of the random integration predominately occurring in the nuclear genome and performed two transformations, distinguished by the promoter, 5'UTR, reporter gene, 3'UTR and the presence of an intron (see supplemented 'DataSheet 1_v1' for GenBank files of full annotated plasmid sequences); (i) the p_{PSAD} :fd-hyd cassette contains an optimized Hydrogenase (HYDA1) enzyme fused to an N-terminal Ferredoxin gene under the control of the native Photosystem I reaction center subunit II (p_{PSAD}) regulatory sequences. It was transformed into a mutant strain of *C. reinhardtii* which shows a negligible production of gaseous hydrogen (*the hyda*₁ and *hyda*₂ double-mutant strain) in order to screen for colonies that restored their ability to produce hydrogen. (ii) the $p_{HSP70A/RBCS2}$:gLuc

plasmid, which is comprised of an optimized Gaussia-Luciferase gene under the control of the native Heat-Shock Protein 70A promoter fused to the promoter of the Rubisco Small subunit 2 (RBCS2), and contains the first intron of the RBCS2 gene. It was transformed into the wild-type CC-124 strain (**Figure 1A**, *Algal Cultures, Growth Conditions and Chlorophyll Determination* and *C. reinhardtii Nuclear Transformation*, the ratio of plasmid to cells used is $\frac{1 \mu\text{g}}{10,000 \text{ cells}}$). For each construct, we obtained 100 colonies for further analysis. To screen for hydrogen-producing colonies, we used an engineered *Rhodobacter capsulatus* that expresses GFP in the presence of gaseous hydrogen (Wecker et al., 2011; Wecker and Ghirardi, 2014) (**Figure 1B**, *Rhodobacter capsulatus High Throughput Screen*). To screen for gLuc positive colonies, we used a Luciferase *in vivo* assay, as described in (Shao and Bock, 2008) (**Figure 1C**, *Luciferase In Vivo Assay*). A total of 33 positive clones (17 fd-hyd and 16 gLuc, ~16% out of the initial collected pool) were collected for further analysis.

DNA Copy-Number and mRNA Level Quantifications

To quantify the number of transgenes integration events that occurred in each of the fd-hyd and the gLuc expressing clones, we extracted total DNA and performed digestion using the restriction enzyme *ScaI*, in order to exclude the possibility that tandem target duplications or targets found in the same chromosome would be encapsulated by the same droplet (*Nucleic Acids Extraction*, the cut site was located outside the amplicons). Next, we used the digested DNA template of each clone for two independent ddPCR reactions using the EvaGreen Supermix; the first reaction amplified an amplicon within the target reporter gene (fd-hyd or gLuc) and the second reaction amplified an amplicon within a reference gene, known to harbor a single copy in the *C. reinhardtii* genome (*CBLP*) (Schloss, 1990). The number of target molecules per μ l ('concentration', retrieved by QuantaSoft™), obtained for each target gene was divided by the concentration obtained for the reference gene (**Figure 2A**); this ratio provided a non-integer number, which was further rounded to obtain the target's DNA copy-number (*Droplet Digital PCR*). Importantly, we did not detect major differences in the sCNV distribution between the fd-hyd and the gLuc expressing clones (**Figure 2B**), thus excluding the possibility that the sequence composition of the coding-sequence or the regulatory elements affects the integration efficiency in our conditions. Interestingly, we observed that the most frequent number of integration events is two (45.5%), whereas one integration site was found in 33.3% of all clones (**Figure 2C**). Additionally, 18.1% and 3.1% of clones contained three and four target copies, respectively (**Figure 2C**).

To quantify the mRNA levels for the fd-hyd clones we anaerobically incubated the cultures (in order to induce the fd-hyd expression for the later Hydrogenase activity assay, which was sampled from the same cultures), extracted total RNA and used a similar ddPCR analysis (without template restriction), normalized by the same reference house-keeping gene *CBLP* (Schloss, 1990) (*Nucleic Acids Extraction, Droplet Digital PCR, fd-hyd Induction*, **Figure 3A**). The same procedure was carried out on the gLuc clones, without anaerobic incubation (**Figure 3B**). Although our

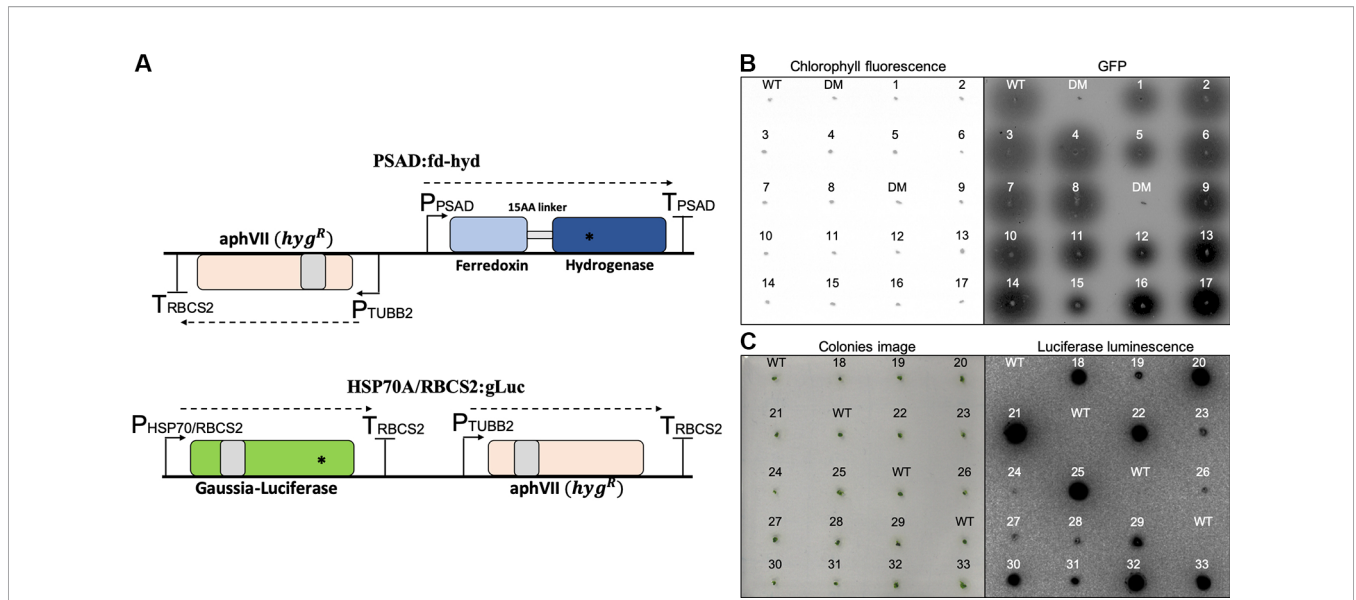


FIGURE 1 | Construction and selection of positive transformants expressing the reporter genes. **(A)** Plasmid maps of the fd-hyd (upper) and the gLuc (bottom) reporter genes. The fd-hyd transgene was cloned into a modified pSL18 vector (the *parO* was replaced by the *hyg^R*) under the control of the PSAD regulatory sequences. The gLuc transgene was cloned into pChlamy_1 plasmid (*hyg^R*) under the control of the HSP70A/RBCS2 promoters and the RBCS2 terminator. Gray boxes represent the first intron of the RBCS2 within the coding-sequences. Asterisks (*) represent the ddPCR amplicon loci of fd-hyd (51bp) and gLuc (85 bp). The cut site of *Scal* is found outside the amplicon area (not shown). **(B)** *Rhodobacter* assay for selected hydrogen producing colonies. fd-hyd colonies were overlaid with engineered H₂-sensing *R. capsulatus* and were scanned using a 473 nm laser for excitation whereas 510 nm or 665 nm filters were used for quantifying GFP fluorescence (right) and chlorophyll density (left), respectively. **(C)** *in-vivo* Luciferase assay for the gLuc selected colonies; gLuc colonies were incubated for 10 min in dark, and then 10 μl of 0.01 mM Coelenterazine were dropped onto each colony. The left panels refer to either the detection of the colonies chlorophyll **(A)** or the photographs of colonies prior to luminescence imaging **(B)**. The right panels refer to GFP fluorescence emitted by *R. capsulatus* **(A)** or luminescence emitted by gLuc **(B)**. WT = CC-124 wild-type strain (which naturally produces hydrogen), DM = *hyda*_{1,2} double-mutant strain (does not produce hydrogen). The numbers refer to clone indices.

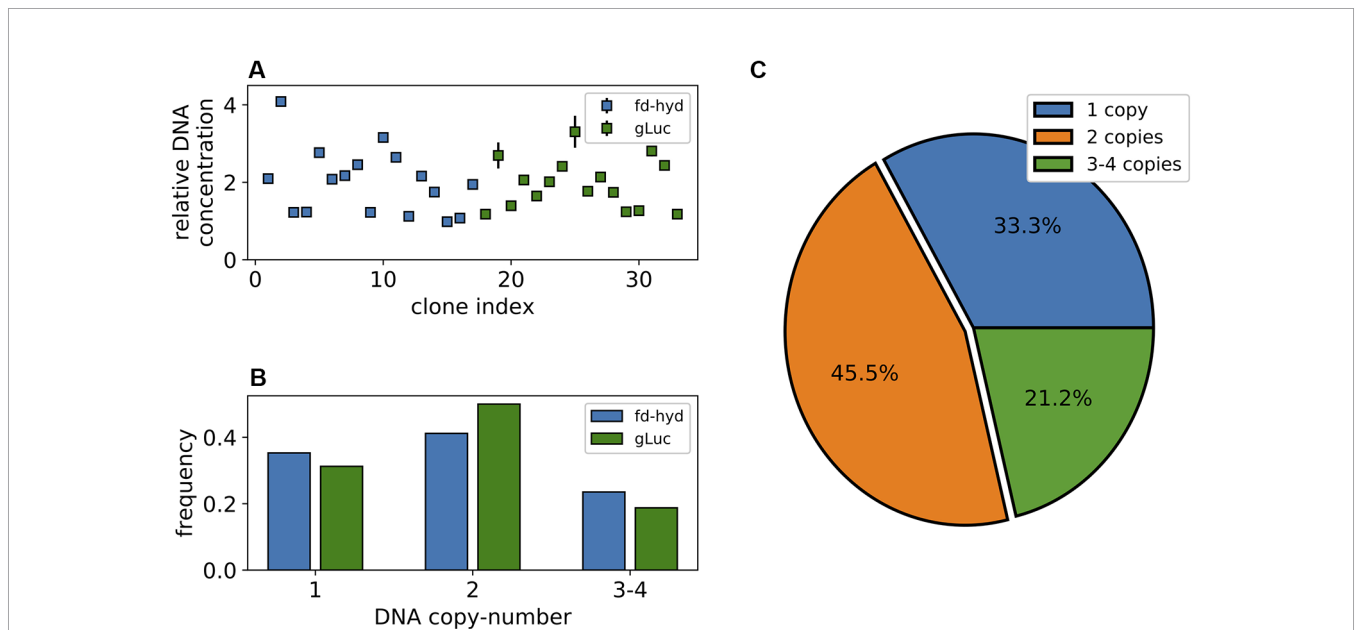


FIGURE 2 | Synthetic copy-number variation of the fd-hyd and the gLuc transgenic clones. **(A)** Transgenes DNA copy-number in 33 nucleus-transformed clones. Copy-number values are calculated as the ratio of the target's concentration to the reference gene (*CBLP*) concentration. The error bars represent the maximum and minimum values of Poisson distribution for the 95% confidence interval, as given by the QuantaSoft™ software. **(B)** The ratio of clones found in each copy-number group separated by the reporter gene. **(C)** Distribution of transgenes' copy-number in all clones.

results show that the gLuc transcript levels are generally lower than the fd-hyd mRNA abundance, we observed a significant correlation between the raw DNA copy-number values (the non-integer numbers received by the ratio of target-to-reference concentration) to the mRNA levels in each reporter separately (**Figures 3C, D** and *Pearson Correlation*; Pearson's rhos are 0.56, $p = 0.02$ and 0.55, $p = 0.027$ for the fd-hyd and the gLuc clones,

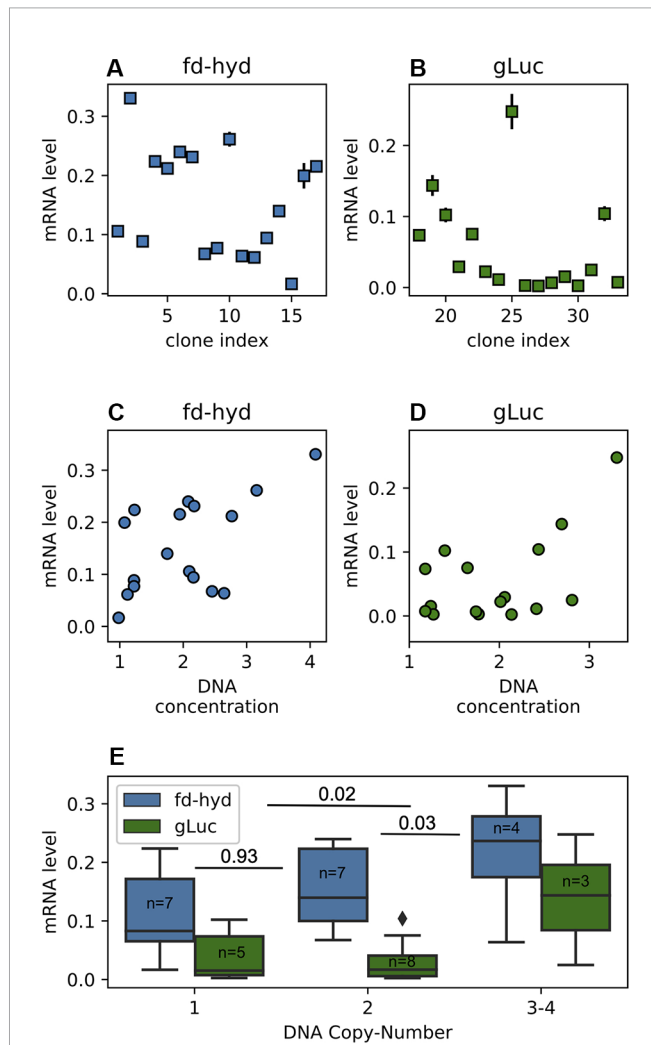


FIGURE 3 | Transgenes mRNA abundance in the fd-hyd and the gLuc clones. **(A, B)** Transgenes relative mRNA abundance in the fd-hyd and the gLuc clones, respectively. The mRNA levels are calculated as the ratio of the target's concentration to the reference gene (*CBLP*) concentration. The error bars represent the maximum and minimum values of Poisson distribution for the 95% confidence interval, as given by the QuantaSoft™ software. **(C, D)** The effect of the relative raw DNA concentration (non-integer values) on transcription levels in the fd-hyd and the gLuc clones, respectively. Pearson correlation coefficients are 0.56, $p = 0.02$ and 0.55, $p = 0.027$ for the fd-hyd and the gLuc clones, respectively. **(E)** The effect of the DNA copy-number on transcription levels. P-values were calculated using two-way ANOVA post-hoc test, and are shown between each copy-number group. The line within each box represents the median. The number of samples for each group is indicated within each box. The copy-number group "3-4" also contains one clone that harbors a copy-number of four (fd-hyd clone #2).

respectively), indicating the positive effect of a transgene copy-number on its transcription level. Moreover, by grouping the mRNA levels to their cognate copy-number values (and by that reducing the variation created by the position effect), we observed a significant difference in transcription between the copy-number groups both in the fd-hyd and the gLuc clones (*Two-Way ANOVA*; 2-way ANOVA $p = 0.002$). Subsequently, we detected a significant separation between two-copies to \geq three-copies groups, whereas the difference between one-copy to two-copies did not yield a significant difference (**Figure 3E** and *Tukey's Range Test*; Tukey's range test $p = 0.03$ and $p = 0.93$, respectively).

Quantification of Protein Activity

To measure the fd-hyd active protein abundance we used the MV quantification assay (*Hydrogenase Activity Assay by Methyl-Viologen*) on three biological repeats (**Table S1**). To measure the gLuc reporter activity, we plated the colonies onto three independent plates while shuffling positions (**Figure S1, Table S1**) and performed a Luciferase *in-vivo* assay (Mayfield and Schultz, 2004; Rasala et al., 2011). To calculate the relative protein activity we used the image analysis software CFQuant (<https://www.energylabtau.com/cfquant>) (Dafni et al., 2019), and obtained the values of "Area" (i.e., sum of pixels) divided by the colony size. In order to reduce the variation between plates even further, these values were divided to the highest value within the same plate; this also allowed us to retrieve a normalized quantity which can further be averaged and compared between plates. We observed a strong positive correlation between the mRNA and protein activity in both constructs (**Figures 4A, B**, Pearson rhos are 0.55, $p = 0.023$ and 0.69, $p = 0.003$ for the fd-hyd and the gLuc clones, respectively). Finally, the direct correlation of the raw DNA values to protein activity levels is significant in the fd-hyd clones, whereas in the gLuc clones this effect is weak (**Figure S2**, Pearson rhos are 0.61, $p = 0.009$, and 0.11, $p = 0.67$ for the fd-hyd and the gLuc clones, respectively).

DISCUSSION

In this work, we suggest that in addition to already known factors (e.g., position-effects, epigenetic-silencing, location in 3D genomic organization, chromatin state, local DNA folding, interruption of native genes), the number of integration sites for a synthetic gene transformed into the nucleus of *C. reinhardtii* is yet another important feature that affects its transcription level and may contribute to the expression-heterogeneity between transgenic clones. By accurately measuring the number of integration sites of two different reporter genes in 33 nucleus-transformed clones, we revealed that the transformed transgene is usually integrated within one to two loci ($\sim 79\%$), whereas in most cases it is inserted two times (**Figure 2C**, based on the plasmid-to-cells ratio of 1:10,000). We observed that the two constructs used in this study show roughly the same distribution of DNA copy-number in the genome, even though they vary in the type of gene, sequence composition, presence of introns and the regulatory elements (**Figure 2B**). Importantly, because fragmented plasmids (that may carry the resistance gene solely) may potentially

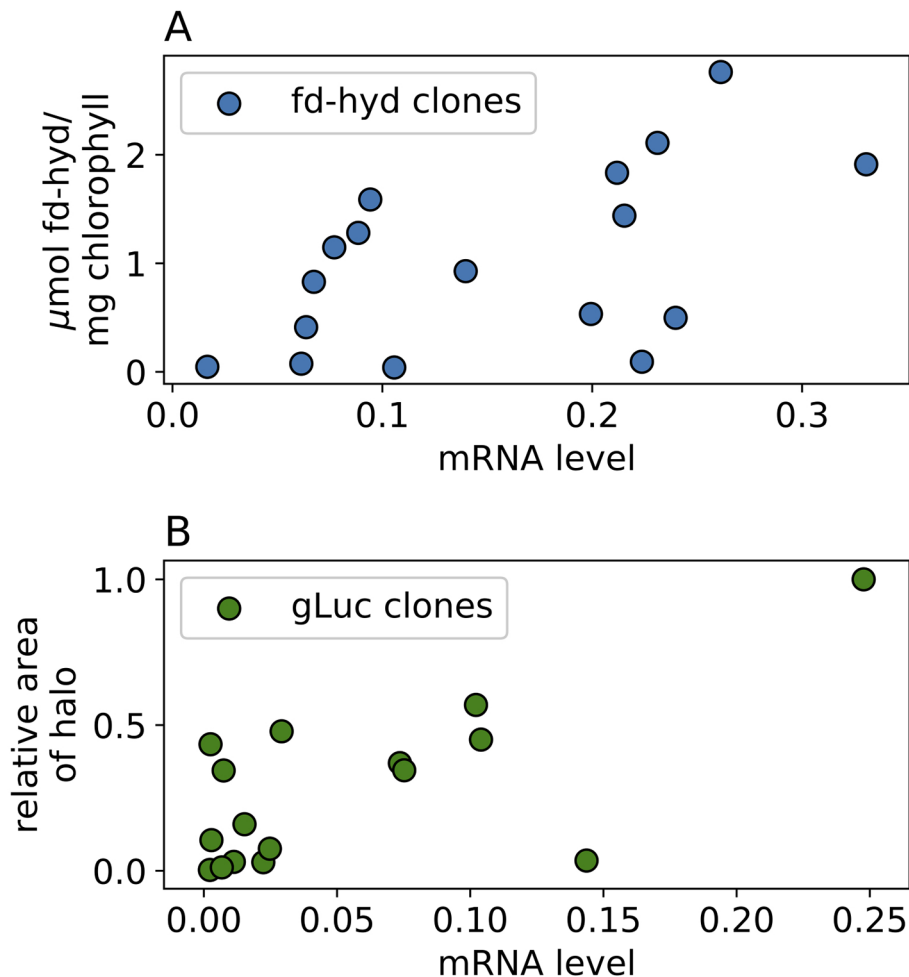


FIGURE 4 | The effect of mRNA levels on protein abundance in the fd-hyd and the gLuc clones. **(A)** the fd-hyd protein activity was measured using MV. The values are $\frac{\mu\text{mol fd-hyd}}{\text{mg chlorophyll}} \times 10^6$ **(B)** gLuc protein values were calculated using CFQuant's "Area" value normalized to the colony size (Dafni et al., 2019). Pearson rhos are 0.55, $p = 0.023$ and 0.69, $p = 0.003$ for the fd-hyd and the gLuc clones, respectively. All values show mean of three biological repeats.

integrate the genome at the time of transformation, we chose to assess the copy-number based on amplicons that are found within the reporter transgene sequences themselves.

To understand the effect of transgene copy-number on transcript levels, we measured the mRNA abundance in the transgenic clones. We were able to show that transgene DNA copy-number positively correlates with its conjugated mRNA levels, with the major effect of three and four copies (**Figures 3C–E**). However, it is critical to note that some clones with multiple integration sites exhibit low transcription levels, and vice-versa (i.e., some clones with a single copy-number show relatively high mRNA levels, **Figures 3C, D**); these exceptions may be explained by other factors (in which their degree varies between transgenes) that can potentially affect transcription; for example (i) the "position-effect", (ii) gene-silencing, (iii) integration of fragmented plasmids, and/or (iv) antagonistic effects of multiple insertion sites. Importantly, these "outliers"

seem to be more frequent in the gLuc clones, a fact that together with their global reduction in mRNA levels (**Figures 3A, B**) emphasizes the notion that the nature of the sequence (e.g., regulatory elements, presence of introns, CDS properties) does affect transcription features. In order to reduce the effect (or "noise") of the above-mentioned factors we presented the data as averages of level of expression, separated by the copy-number groups (**Figure 3E**). Based on our experimental conditions, three integration events had the highest impact on transcription. According to this, one can estimate that multiple DNA copies of the transgene do not synergistically affect transcription (as a major part of transgene copies would not be expressed at all); alternatively, it can be hypothesized that the higher the transgene's copy-number is, the chances of at least one cassette to integrate into a highly expressed DNA region are increased. With that, it is tempting to use sCNV analysis in order to screen for highly expressing clones in *C. reinhardtii* (when no efficient

screening method exists) by searching for clones that harbor at least three copies of the integrated transgene.

Finally, we measured the recombinant protein abundance (as a function of protein activity) in our library (Figures 4A, B); the results align with earlier evidences, where the correlation between mRNA to protein levels in the nuclear genome of *C. reinhardtii* is high (Baier et al., 2018; Weiner et al., 2018). Yet, several lines show notable exceptions from this rule (e.g., clones with high mRNA levels show low/negligible levels of protein); these results may be explained by either post-transcriptional mechanisms of gene-silencing, and/or deletions or mutations at the synthetic gene regulatory elements (which may reduce translation efficiency), or at the CDS region (which may produce non-mature mRNAs, or yield a non-functional protein). Lastly, the direct effect of DNA copy-number on protein levels in the fd-hyd clones is significantly sound (Figure S2A); however, the same analysis yielded poor results on the gLuc clones (Figure S2B). As mentioned above, as the mRNA levels of the gLuc lines are less affected by their cognate DNA copy-number, let alone the protein levels - which are subject to more regulations (e.g., post-transcriptional and post-translational modifications).

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/Supplementary Material.

REFERENCES

- Baier, T., Wichmann, J., Kruse, O., and Lauersen, K. J. (2018). Intron-containing algal transgenes mediate efficient recombinant gene expression in the green microalga *Chlamydomonas reinhardtii*. *Nucleic Acids Res.* 46, 6909–6919. doi: 10.1093/nar/gky532
- Barahimipour, R., Strenkert, D., Neupert, J., Schroda, M., Merchant, S. S., and Bock, R. (2015). Dissecting the contributions of GC content and codon usage to gene expression in the model alga *Chlamydomonas reinhardtii*. *Plant J.* 84, 704–717. doi: 10.1111/tbj.13033
- Barahimipour, R., Neupert, J., and Bock, R. (2016). Efficient expression of nuclear transgenes in the green alga *Chlamydomonas*: synthesis of an HIV antigen and development of a new selectable marker. *Plant Mol. Biol.* 90, 403–418. doi: 10.1007/s11103-015-0425-8
- Collier, R., Dasgupta, K., Xing, Y.-P., Hernandez, B. T., Shao, M., Rohozinski, D., et al. (2017). Accurate measurement of transgene copy number in crop plants using droplet digital PCR. *Plant J.* 90, 1014–1025. doi: 10.1111/tbj.13517
- Dafni, E., Weiner, I., Shahar, N., Tuller, T., and Yacoby, I. (2019). Image-processing software for high-throughput quantification of colony luminescence. *mSphere* 4, e00676–18. doi: 10.1128/msphere.00676-18
- De Block, M. (1993). The cell biology of plant transformation: current state, problems, prospects and the implications for the plant breeding. *Euphytica*. 71, 1–4. doi: 10.1007/BF00023461
- Demeke, T., Gräfenhan, T., Holigroski, M., Fernando, U., Bamforth, J., and Lee, S. J. (2014). Assessment of droplet digital PCR for absolute quantification of genetically engineered OXY235 canola and DP305423 soybean samples. *Food Control*. 46, 470–474. doi: 10.1016/j.foodcont.2014.06.018
- Eilenberg, H., Weiner, I., Ben-Zvi, O., Pundak, C., Marmari, A., Liran, O., et al. (2016). The dual effect of a ferredoxin-hydrogenase fusion protein in vivo: Successful divergence of the photosynthetic electron flux towards hydrogen production and elevated oxygen tolerance. *Biotechnol. Biofuels* 9, 182. doi: 10.1186/s13068-016-0601-3

AUTHOR CONTRIBUTIONS

NS and IY designed the research. NS, SL, IW, TE, and YF performed the experimental procedures. NS, IW, and ED performed the computational procedures. NS, IW, TT and IY wrote the paper.

FUNDING

The single source of this research is Israel Science foundation grant number 1646/16.

ACKNOWLEDGMENTS

We would like to thank Professor Matthew Posewitz for the *hyda*_{1,2} double knockout mutant, Professors Matt Wecker and Maria Ghirardi for the H₂-sensing *R. capsulatus*, and Dr. Hila Kobo for ddPCR consultation.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2019.01784/full#supplementary-material>

- Fischer, N., and Rochaix, J. D. (2001). The flanking regions of PsaD drive efficient gene expression in the nucleus of the green alga *Chlamydomonas reinhardtii*. *Mol. Genet. Genomics* 265, 888–894. doi: 10.1007/s004380100485
- Greiner, A., Kelterborn, S., Evers, H., Kreimer, G., Sizova, I., and Hegemann, P. (2017). Targeting of photoreceptor genes in *Chlamydomonas reinhardtii* via zinc-finger nucleases and CRISPR/Cas9. *Plant Cell*. 29, 2498–2518. doi: 10.1105/tpc.17.00659
- Hindson, B. J., Ness, K. D., Masquelier, D. A., Belgrader, P., Heredia, N. J., Makarewicz, A. J., et al. (2011). High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Anal. Chem.* 83, 8604–8610. doi: 10.1021/ac202028g
- Hobbs, S. L. A., Kpodar, P., and DeLong, C. M. O. (1990). The effect of T-DNA copy number, position and methylation on reporter gene expression in tobacco transformants. *Plant Mol. Biol.* 15, 851–864. doi: 10.1007/BF00039425
- Hobbs, S. L., Warkentin, T. D., and DeLong, C. M. (1993). Transgene copy number can be positively or negatively associated with transgene expression. *Plant Mol. Biol.* 21, 17–26. doi: 10.1007/BF00039614
- Jinkerson, R. E., and Jonikas, M. C. (2015). Molecular techniques to interrogate and edit the *Chlamydomonas* nuclear genome. *Plant J.* 82, 393–412. doi: 10.1111/tbj.12801
- Jorgensen, R. A., Cluster, P. D., English, J., Que, Q., and Napoli, C. A. (1996). Chalcone synthase cosuppression phenotypes in petunia flowers: Comparison of sense vs. antisense constructs and single-copy vs. complex T-DNA sequences. *Plant Mol. Biol.* 31, 957–973. doi: 10.1007/BF00040715
- Ku, M. S. B., Agarie, S., Nomura, M., Fukayama, H., Tsuchida, H., Ono, K., et al. (1999). High-level expression of maize phosphoenolpyruvate carboxylase in transgenic rice plants. *Nat. Biotechnol.* 17, 76. doi: 10.1038/5256
- Lauersen, K. J., Baier, T., Wichmann, J., Wördenweber, R., Mussnug, J. H., Hübner, W., et al. (2016). Efficient phototrophic production of a high-value sesquiterpenoid from the eukaryotic microalga *Chlamydomonas reinhardtii*. *Metab. Eng.* 38, 331–343. doi: 10.1016/j.ymben.2016.07.013

- León, R., and Fernández, E. (2007). Nuclear transformation of eukaryotic microalgae: historical overview, achievements and problems. *Adv. Exp. Med. Biol.* 616, 1–11. doi: 10.1007/978-0-387-75532-8_1
- Lodha, M., Schulz-Raffelt, M., and Schroda, M. (2008). A new assay for promoter analysis in *Chlamydomonas* reveals roles for heat shock elements and the TATA box in HSP70A promoter-mediated activation of transgene expression. *Eukaryot. Cell.* 7, 172–176. doi: 10.1128/EC.00055-07
- Malhotra, D., and Sebat, J. (2012). CNVs: Harbingers of a rare variant revolution in psychiatric genetics. *Cell.* 148, 1223–1241. doi: 10.1016/j.cell.2012.02.039
- Mayfield, S. P., and Schultz, J. (2004). Development of a luciferase reporter gene, luxCt, for *Chlamydomonas reinhardtii* chloroplast. *Plant J.* 37, 449–458. doi: 10.1046/j.1365-313X.2003.01965.x
- Meuser, J. E., D'Adamo, S., Jinkerson, R. E., Mus, F., Yang, W., Ghirardi, M. L., et al. (2012). Genetic disruption of both *Chlamydomonas reinhardtii* [FeFe]-hydrogenases: insight into the role of HYDA2 in H₂ production. *Biochem. Biophys. Res. Commun.* 417, 704–709. doi: 10.1016/j.bbrc.2011.12.002
- Neupert, J., Karcher, D., and Bock, R. (2009). Generation of *Chlamydomonas* strains that efficiently express nuclear transgenes. *Plant J.* 57, 1140–1150. doi: 10.1111/j.1365-313X.2008.03746.x
- Nour-Eldin, H. H., Specht, E. A., Ostrand, J., Hoang, K. T. D., Karunanithi, P. S., and Mayfield, S. P. (2018). High-throughput system for quantifying and characterizing homologous recombination in *Chlamydomonas reinhardtii*. *Algal Res.* 31, 167–172. doi: 10.1016/j.algal.2018.02.005
- Pinheiro, L. B., Coleman, V. A., Hindson, C. M., Herrmann, J., Hindson, B. J., Bhat, S., et al. (2012). Evaluation of a droplet digital polymerase chain reaction format for DNA copy number quantification. *Anal. Chem.* 84, 1003–1011. doi: 10.1021/ac202578x
- Que, Q., Wang, H. Y., English, J. J., and Jorgensen, R. A. (1997). The frequency and degree of cosuppression by sense chalcone synthase transgenes are dependent on transgene promoter strength and are reduced by premature nonsense codons in the transgene coding sequence. *Plant Cell.* 9, 1357–1368. doi: 10.1105/tpc.9.8.1357
- Rasala, B. A., Muto, M., Sullivan, J., and Mayfield, S. P. (2011). Improved heterologous protein expression in the chloroplast of *Chlamydomonas reinhardtii* through promoter and 5' untranslated region optimization. *Plant Biotechnol. J.* 9, 674–683. doi: 10.1111/j.1467-7652.2011.00620.x
- Redon, R., Ishikawa, S., Fitch, K. R., Feuk, L., Perry, G. H., Andrews, T. D., et al. (2006). Global variation in copy number in the human genome. *Nature.* 444, 444–454. doi: 10.1038/nature05329
- Schloss, J. A. (1990). A *Chlamydomonas* gene encodes a G protein β subunit-like polypeptide. *MGG Mol. Gen. Genet.* 221, 443–452. doi: 10.1007/BF00259410
- Schubert, D., Lechtenberg, B., Forsbach, A., Gils, M., Bahadur, S., and Schmidt, R. (2004). Silencing in Arabidopsis T-DNA transformants: the predominant role of a gene-specific RNA sensing mechanism versus position effects. *Plant Cell.* 16, 2561–2572. doi: 10.1105/tpc.104.024547
- Shao, N., and Bock, R. (2008). A codon-optimized luciferase from *Gaussia princeps* facilitates the in vivo monitoring of gene expression in the model alga *Chlamydomonas reinhardtii*. *Curr. Genet.* 53, 381–388. doi: 10.1007/s00294-008-0189-7
- Shao, N., Krieger-Liszkay, A., Schroda, M., and Beck, C. F. (2007). A reporter system for the individual detection of hydrogen peroxide and singlet oxygen: its use for the assay of reactive oxygen species produced in vivo. *Plant J.* 50, 475–487. doi: 10.1111/j.1365-313X.2007.03065.x
- Shaver, S., Armando Casas-Mollano, J., Cerny, R. L., Cerutti, H. D., and Armando, J. (2010). Origin of the polycomb repressive complex 2 and gene silencing by an E(z) homolog in the unicellular alga *Chlamydomonas*. *Epigenetics* 54, 301–312. doi: 10.4161/epi.5.4.11608
- Shin, S. E., Lim, J. M., Koh, H. G., Kim, E. K., Kang, N. K., Jeon, S., et al. (2016). CRISPR/Cas9-induced knockout and knock-in mutations in *Chlamydomonas reinhardtii*. *Sci. Rep.* 6, 27810. doi: 10.1038/srep27810
- Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98, 503–517. doi: 10.1016/S0022-2836(75)80083-0
- Stefano, B., Patrizia, B., Cerboneschi, M., and Massimo, G. (2016). Inverse PCR and Quantitative PCR as Alternative Methods to Southern Blotting Analysis to Assess Transgene Copy Number and Characterize the Integration Site in Transgenic Woody Plants. *Biochem. Genet.* 54, 291–305. doi: 10.1007/s10528-016-9719-z
- Stockhaus, J., Eckes, P., Blau, A., Schell, J., and Willmitzer, L. (1987). Organ-specific and dosage-dependent expression of a leaf/stem specific gene from potato after tagging and transfer into potato and tobacco plants. *Nucleic Acids Res.* 15, 3479–3491. doi: 10.1093/nar/15.8.3479
- Wang, Y., Xiong, G., Hu, J., Jiang, L., Yu, H., Xu, J., et al. (2015). Copy number variation at the GL7 locus contributes to grain size diversity in rice. *Nat. Genet.* 47, 944–948. doi: 10.1038/ng.3346
- Wecker, M. S. A., and Ghirardi, M. L. (2014). High-throughput biosensor discriminates between different algal H₂-photoproducing strains. *Biotechnol. Bioeng.* 111, 1332–1340. doi: 10.1002/bit.25206
- Wecker, M. S. A., Meuser, J. E., Posewitz, M. C., and Ghirardi, M. L. (2011). Design of a new biosensor for algal H₂ production based on the H₂-sensing system of *Rhodospirillum rubrum*. *Int. J. Hydrogen Energy* 36, 11229–11237. doi: 10.1016/j.ijhydene.2011.05.121
- Weiner, I., Atar, S., Schweitzer, S., Eilenberg, H., Feldman, Y., Avitan, M., et al. (2018). Enhancing heterologous expression in *Chlamydomonas reinhardtii* by transcript sequence optimization. *Plant J.* 94, 22–31. doi: 10.1111/tpj.13836
- Wichmann, J., Baier, T., Wentnagel, E., Lauersen, K. J., and Kruse, O. (2018). Tailored carbon partitioning for phototrophic production of (E)- α -bisabolene from the green microalga *Chlamydomonas reinhardtii*. *Metab. Eng.* 45, 211–222. doi: 10.1016/j.ymben.2017.12.010
- Wu-scharf, D., and Cerutti, H. D. (2000). Transgene and transposon silencing in *Chlamydomonas reinhardtii* by a DEAH-Box RNA helicase transgene and transposon silencing in *Chlamydomonas reinhardtii* by a DEAH-Box RNA Helicase. *Science* (80-), 1159–1162. doi: 10.1126/science.290.5494.1159
- Yacoby, I., Pochekailov, S., Toporik, H., Ghirardi, M. L., King, P. W., and Zhang, S. (2011). Photosynthetic electron partitioning between [FeFe]-hydrogenase and ferredoxin:NADP⁺-oxidoreductase (FNR) enzymes in vitro. *Proc. Natl. Acad. Sci.* 108, 9396–9401. doi: 10.1073/pnas.1103659108
- Yamano, T., Iguchi, H., and Fukuzawa, H. (2013). Rapid transformation of *Chlamydomonas reinhardtii* without cell-wall removal. *J. Biosci. Bioeng.* 115, 691–694. doi: 10.1016/j.jbiosc.2012.12.020
- Zhang, R., Patena, W., Armbruster, U., Gang, S. S., Blum, S. R., and Jonikas, M. C. (2014). High-throughput genotyping of green algal mutants reveals random distribution of mutagenic insertion sites and endonucleolytic cleavage of transforming DNA. *Plant Cell.* 26, 1398–1409. doi: 10.1105/tpc.114.124099
- Zhang, Y., Liu, Y., Zhang, J., Wang, G., Wang, J., and Liu, Y. (2015). Assessment of transgene copy number and zygosity of transgenic maize overexpressing Cry1Ie gene with SYBRi₂ Green qRT-PCR. *Vitr. Cell. Dev. Biol. - Plant* 51, 125–134. doi: 10.1007/s11627-014-9658-5

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Shahar, Landman, Weiner, Elman, Dafni, Feldman, Tuller and Yacoby. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.