Gene Reports 11 (2018): 235-238

**DOI:** https://doi.org/10.1016/j.genrep.2018.04.005

**Cytogenetic Identity: a new parameter for estimating whole-genome differences**

Mariella Nieddu[a], Renato Robledo[a*], Roberto Mezzanotte[a], Giuseppe Massimo Vacca[b], Miguel Pita[c]

[a] Department of Biomedical Sciences, University of Cagliari, Italy
[b] Department of Animal Biology, University of Sassari, Italy
[c] Department of Biology, Universidad Autónoma de Madrid, Spain

* Corresponding author

Address: Department of Biomedical Sciences, University of Cagliari, Cittadella Universitaria, Monserrato, 09042, Italy
Email: rrobledo@unica.it

mnieddu@unica.it
mezzanotte@unica.it
gmvacca@uniss.it
miguel.pita@uam.es

**Abstract**

In this paper, we introduce a novel parameter, called Cytogenetic Identity, to describe differences and similarities between genomes. Using Whole Comparative Genomic Hybridization plus Digital Image Analysis, we present a new methodology that employs the whole genome, including highly repeated DNA sequences, to provide a general picture about the differences between individuals of the same or different species. The proposed approach has a great potential in many different fields of research, like evolution, ecology, phylogenesis, etc. In the present study, we applied Cytogenetic Identity to establish a quantitative degree of divergence between different goat breeds. Advantages as well as disadvantages of the new parameter are discussed.

**Keywords:** Whole Comparative Genomic Hybridization, Satellite DNA, Cytogenetic Identity, Species, Chromosomes.

## 1. Introduction

Currently observed biodiversity has been shaped by millions of years of evolution and, more recently, by human intervention. Biodiversity can be described by analyzing differences occurring within and between species. The ability to establish resemblances and differences between individuals of the same and/or different species is of a paramount importance for researchers in many different fields. From the evolutionary interest in determining the degree of divergence of species, to the necessity of precisely describing the polymorphisms that are responsible for diseases, or to manage threatened and endangered species.

Initially, species were compared by morphological criteria; subsequently, with the development of molecular biology, degree of divergence was based mainly on protein and DNA analysis. In particular, DNA sequencing of a 658 bp fragment of the mitochondrial gene cytochrome *c* oxidase I (COI) allowed a DNA "barcoding" system (Hebert et al 2003). In the near future, next generation sequencing (NGS) may allow a taxonomy based on entire genome analysis (Ellegren 2008).

However, since DNA analysis is focused on coding sequences, even whole sequencing studies will miss non-coding regions, like chromosome centromeres. Centromeres are the loci responsible for the correct segregation of chromosomes during mitosis and meiosis, and they are contained within regions of highly repetitive sequences, called satellite DNA. Centromeres are stably inherited: however, being non-coding, their DNA sequences are not under evolutionary constraint and show a higher mutation rate. Therefore, centromeric sequences, as well as other highly repeated DNAs, are of particular interest for studies on evolution, since they can depict mutation accumulation proportional to the divergence rate. Indeed, centromere evolution has been proposed to be a key factor in speciation (Henikoff et al 2001).

In this paper, we propose a new tool to compare genomes of individuals of the same or different species, named Cytogenetic Identity (CI). This new parameter allows to include the whole genomes of the involved individuals in the comparison study, providing a general picture. We define and estimate Cytogenetic Identity between individuals as a reflection of the degree of divergence they have experienced throughout evolution. Our results provide a novel insight in the differences between genomes that, although less precise than molecular, bears into account all the genome sequences, including those with a higher mutation rate.

To estimate Cytogenetic Identity, we use Whole-Comparative Genomic Hybridization (W-CGH), a technique that allows detecting in a single Fluorescent *in situ* Hybridization (FISH) protocol all the chromosome differences between too compared genomes (Pita et al., 2003). The technique has proven its reliability in different species (Pita et al., 2008, 2009) and, in this new approach complemented with Digital Image Analysis (DIA), it allow us to obtain quantitative information. This new use permits the estimation of Cytogenetic Identity offering useful application in the comparison of closely related individuals. In this paper, we analyzed individuals belonging to different goat breeds, as well as closely related species, since they provide us with a wide range of breeds to explore the accuracy of the approach.

## 2. Materials and Methods

### *2.1. Samples*

Blood samples were collected from *Capra hircus* (Sarda breed, Maltese breed and Murciana breed), *Ovis aries* and *Bos taurus*. DNA and metaphase chromosomes were obtained with standard methods. ARRIVE guidelines have been followed.

### *2.2 Probe labeling*

DNA labeling with direct fluorochromes was performed using a Nick Translation kit (Enzo Life Sciences). One μg of each DNA was independently labeled with 0.3 mM Green-496 dUTP or Orange-552 dUTP (Enzo Life Sciences). Probe fragments were checked on 1% agarose gel to be similar in size and in the range of 600-2000 bp. Probes were then precipitated overnight with ethanol, centrifuged at full speed and supernatant was discarded. After complete air drying, probes were dissolved in hybridization buffer containing 50% (vol/vol) formamide, 10% (w/vol) dextran sulfate in 2x SSC, at pH 7, to a final concentration of 20 ng/μl.

### *2.3. W-CGH (Whole-Comparative Genomic Hybridization)*

Comparative Genomic Hybridization of any two probes was performed as follows: slides were dehydrated in an ethanol series (70%, 85% and 100%), for 3 min each, at -

20ºC. After air drying, slides were denatured in 2x SSC, with 70% formamide, for 2 min at 73ºC, dehydrated and dried again. Mixed probes were prepared adding equimolar concentration of the two labeled genomes to be compared (one probe labeled with Green-496 dUTP and the other probe labeled with Orange-552 dUTP) to a final volume of 15 µl. Mixed probes were denatured for 10 min at 73ºC, chilled on ice for 5 min and applied to the slide. Slides were incubated on wet chamber for 16 h at 37ºC for hybridization. After hybridization, slides were washed in in 2x SSC, with 50% formamide, for 15 min at 42ºC, and in 2x SSC for 8 min at 37ºC. Finally, slides were mounted with anti-fade solution (Vectashield, Vector Laboratories) and counterstained with 4´,6-Diamidino-2-Phenylindole (DAPI) (100 ng/µl).

For each two species, breeds or individuals compared, the same set of two slides was prepared. It consisted of one slide with Sarda-breed cells hybridized with a mixed probe composed of Sarda genome labeled with Green-498 dUTP and an *Alien* genome labeled with Orange-552 dUTP. As a control, a second slide with Sarda-breed cells was hybridized with the inverted mixed probe, i.e. Sarda genome labeled in Red and the *Alien* genome labeled in Green.

*2.4. Image capturing and Digital Image Analysis (DIA)*

Slides were analyzed using a DIA platform based on a Leica DMRB fluorescence microscope (Leica Microsystems) with three independent filters for Green-496 dUTP (Green fluorescence), Orange-552 dUTP (Red fluorescence) and DAPI (Blue fluorescence) detection (I3, Y3, and DAPI, respectively). Images were captured as three independent .tiff files (Green channel, Red channel and Blue channel) employing Leica DFC 350 FX (Leica Microsystems) running in Adobe Photoshop software (Adobe Systems Incorporated). Several images (at least 20) were captured to analyze interphases nuclei for quantitative results, as well as mitotic metaphases for qualitative description of the hybridization.

Adobe Photoshop software was also used to merge the Green, Red and Blue channels to create an RGB image after background subtraction. Also a larger image with several nuclei (5-15) was created assembling all the nuclei images of the same slide to facilitate DIA. Mitotic metaphase images were not employed for DIA since interphase nuclei provide a more homogeneous and individualized material to capture more detailed fluorescence information.

DIA of FISH images was performed employing ImageJ software (https://imagej.nih.gov/ij/). For each RGB image, Blue channel (unspecific DAPI counterstaining) was used to select the area to be measured on each cell (Figure 1). Green and Red Fluorescence on each nucleus was measured, under that area, as Sum of Grey in the range of 0 to 255 (for Green and Red, independently). Area of each cell as the number of pixels was also recorded. Results were exported to Excel to estimate Cytogenetic Identity (CI) from Sum of Grey and Pixels area.

*2.5. Cytogenetic Identity (CI)*

Several parameters were calculated using Sum of Grey (of Green and Red independently) and the Area of each cell (number of pixels). First, Average Grey (in the range of 0-255) was estimated, for each single cell, as the division of the Sum of Grey (for Green and Red independently) and its Area. Afterwards, Average Grey 2 (for Green and Red independently) was estimated as the mean value of all the Average Grey values of the cells on each experiment, and the same was done for the converse experiment (i.e. Sarda-G/*Alien*-R and Sarda-R/*Alien*-G).

Then, we calculated a Sarda Average Grey (SAG) and an *Alien* Average Grey (AAG). For example, to estimate SAG we employed the Average Grey 2 of Green from the experiment Sarda-G/*Alien*-R and the Average Grey 2 of Red from Sarda-R/*Alien*-G. In this same experiment, for AAG we used the Average Grey 2 of G from Sarda-R/*Alien*-G and the Average Grey 2 of Red from the experiment Sarda-G/*Alien*-R. Before estimating SAG and AAG, we confirmed that there were not statistically significant differences between the Average Grey 2 values to be merged (t-Stu; $p>0,05$). Finally, using these parameters, Cytogenetic Identity was estimated for each two compared genomes as 2xAAG/(AAG+SAG) (Figure 2).

This formula provides a reflection of the rate to which to two genomes share regions of homology. That is because from the W-CGH rationale we can deduct that in those areas of the cell nucleus where the two compared genomes are homologous, both genomes in the mixed-probe have equal chances of hybridizing. Therefore those regions in the nuclei will display shared G and R fluorescence. On the contrary, those other regions of the nucleus harboring sequences that the *Alien* genome does not share with the (Sarda) host genome can only be hybridized by the homologous genome in the mixed-probe, exclusively displaying Sarda probe fluorescence. Therefore, host probe

hybridization (Sarda in all these experiments) can be seen in regions where only host is able to hybridize and also in regions that are shared with the *alien* genome. On the contrary, *alien*-probe hybridizes half the regions where it has homology (the other half will display host's hybridization). This rationale provides the required information to estimate CI: that is the *alien* probe signal will represent half of the homology between the host and *alien* genomes. Consequently, CI is estimated using 2x*Alien* (AAG), divided by the total hybridization (AAG + SAG), values always estimated from the measured Grey levels.

## 3. Results

### *3.1. Quantitative results*

To estimate Cytogenetic Identity, Sarda goat genome was compared to several other genomes (of the same and different breed and species). In all cases W-CGH experiments were performed over Sarda goat-cell nuclei.

The following different experiments were set up: Sarda goat genome (from now *Standard*) was compared to another Sarda goat genome (named B), to Maltese goat genome and to Murciana goat genome. In addition, it was compared to Sheep and Cattle genomes as external groups, and to its own genome as a control.

For each comparison a mixed-probe Sarda-G/*Alien*-R probe and a Sarda-R/*Alien*-G probe were hybridized. On each slide 5-10 homogeneously looking cells were measured to obtain the Cytogenetic Identity (Table 1)

| Sarda-Standard vs: | 2xAAG/(AAG+SAG) | CI x 100 |
|---|---|---|
| Goat Sarda B | CI = 1.10 (n=9+10) | 110% |
| Goat Maltese | CI = 1.02 (n=5+6) | 102% |
| Goat Murciana | CI = 0.83 (n=9+8) | 83% |
| Sheep | CI = 0.78 (n=5+8) | 78% |
| Cattle | CI = 0.72 (n=7+11) | 72% |
| Goat Sarda-Standard | CI = 1.09 (n=9) | 109% |

Table 1. Cytogenetic Identities of several genomes compared to Sarda's. Number of studied cells are included in parenthesis. Comparison of Sarda *versus* Sarda B and the control (Sarda-Standard) show values close to 100% indicating the similarity of the studied genomes the reliability of the study.

Our results provide a general perspective of the degree of similarity of the studied genomes compared to that of the Sarda breed. When Sarda breed genome is compared to itself or to another Sarda breed genome, the CI is close to 1 (CI=1.09 and CI=1.10, respectively), probing the reliability of the approach. When compared to Maltese breed genome a similar situation is observed (CI=1.02), since we are analyzing an individual of a very close breed. Some distinguishable differences appear when Sarda breed is compared to Murciana (CI=0.83), a more distant breed. Larger differences were observed when Sarda breed was compared to Sheep (CI=0.78) and Cattle (CI=0.72), two different species.

### *3.2. Qualitative results*

The analysis of the mitotic metaphases provides information of those regions revealing homology and divergence between the compared genomes.
Whereas chromosome arms reveal a yellow-like fluorescence (of the similar hybridization of G and R genomes in the mixed probe), highly repeated DNAs are the regions that reveal the strongest differences and hence the higher degree of divergence. When comparing two Sarda genomes the whole karyotype reveals yellow-like fluorescence, including centromeres. But the more different two breeds or species are, the higher prevalence of Sarda-probe fluorescence we observe (Figure 3). Consistent with quantitative results, centromeres reveal only Sarda-probe fluorescence when Sarda breed was compared to Cattle (in R in Figure 3e).

## 4. Discussion

In the present paper we introduce a novel parameter, Cytogenetic Identity, as a new tool to evaluate the degree of divergence between individuals or species. This new parameter takes into consideration the entire genome, including repetitive, non-coding, sequences

like satellite DNA, which reveal the most remarkable differences, given that euchromatic regions are quite similar among related individuals, breeds or species.

The applications of this new parameter are determined by the degree of resolution: individuals of the same, or too closely related, species will show a CI value close to 100%, as we observed comparing Sarda vs Sarda, or Sarda vs Maltese. Likewise, if the species are too distant, CI may not be particularly informative either: this is because employing W-CGH most of the differences are due to the highly repeated DNA sequences. Indeed, it is expected that in chromosome arms there is always a certain degree of hybridization no matter how distant the species are (Kallioniemi et al., 1992). Therefore, differences between distant and very distant species will be more difficult to observe. However, CI is ideal for different breeds or closely related species, where chromosome arms will show similarities as well as differences detectable in the highly repeated DNAs, which can provide details and divergence times that not even Next Generation Sequencing will be able to reveal. Our approach allows to look into the whole genome in a single easy and affordable experiment based on fluorescence techniques.

In our study, we observed CIs in the range of 72% (Cattle), 78% (Sheep), 83% (Murciana goat), and 100% (Maltese goat). The degree of divergence between Sarda and Murciana highlights the main goal of our approach: the use of highly repeated DNA sequences, commonly overlooked in genomic studies. Qualitative results reveal that when comparing genomes at the cytogenetic level, most of the differences are present at the centromeric regions, since differences in euchromatin are too subtle for the resolution of this technique. There is little doubt that all satellite sequences have an evolutionary history behind them, which could be revealed with our approach. CI turns highly repeated DNAs into interesting sequences. These non-coding regions are of great interest to date differences between genomes since they are the home for a constant mutation rate, and therefore they are able to provide a more accurate divergence rate that coding regions.

Our results are in agreement with the currently proposed phylogeny that places sheep and goat in the same subfamily of *Caprinae*, while cattle belongs the *Bovinae* subfamily. Indeed, CI values are lower when goat is compared to cattle rather than when goat is compared to sheep. However, our results do not provide a rigorous measure of the real differences between these breeds and species, since we just use one individual to test the approach. In future studies, in order to obtain reliable quantitative results of

Cytogenetic Identity, it is essential to use several individual DNAs to represent each breed or species.

Since no previous knowledge of sequence information is required to determine Cytogenetic Identity, its field of applications may include a wide range of species. Particularly those that are closely related and tend to show small differences between individuals under standard approaches; the use of highly repeated sequences allows to detect subtle divergence rates. An additional advantage is that the technique may be employed in those species where chromosomes are very hard to harvest and only interphase nuclei are accessible.

In conclusion, we believe that Cytogenetic Identity may become the method of choice for evaluating degree of divergence in closely related species, or breeds. This new parameter may provide additional information that could be invaluable in solving the still controversial phylogenetic relationship like, for example, the ones involving differents species of the Bovidae family.

## 5. Acknowledgements

The work was supported by Fondo Integrativo per la Ricerca (F.I.R.), provided by the University of Cagliari (to M.N. and R.R.)

## 6. References

Ellegren H (2008) Sequencing goes 454 and takes large-scale genomics into the wild. Mol Ecol 17: 1629-31

Hebert PD, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. Proc Biol Sci 270: 313-21

Henikoff S, Ahmad K, Malik HS (2001) The centromere paradox: stable inheritance with rapidly evolving DNA. Science 293: 1098-102

Kallioniemi A, Kallioniemi OP, Sudar Da, Rutovitz D, Gray JW, Waldman F, Pinkel D (1992) Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. Science 258:818-821.

Pita M, Fernández JL, Gosálvez J (2003) Whole-comparative genomic hybridization (W-CGH): 1. The quick overview of repetitive DNA sequences on a genome. Chromosome Res 11: 673-679

Pita M, Zabal-Aguirre M, Arroyo F, Gosálvez J, López-Fernández C, de la Torre J (2008) *Arcyptera fusca* and *Arcyptera tornosi* repetitive DNA families: whole-comparative genomic hybridization (W-CGH) as a novel approach to the study of satellite DNA libraries. J Evol Biol 21: 352-361

Pita M, Gosálvez J, Gosálbez A, Nieddu M, López-Fernández C, Mezzanotte R (2009) A highly conserved pericentromeric domain in human and gorilla chromosomes. Cytogenet Genome Res 126: 253-258

**Figure captions**

Figure 1. From left to right: Example of a nucleus to be measured, selection of the region of interest using the DAPI unspecific counterstaining and mask of the region under which ImageJ software is automatically measuring Green and Red Grey levels.

Figure 2. Cells from an experiment (Sarda-G/Alien-R) (above) and from the converse experiment (Sarda-R/Alien-G) (below). Estimation of the SAG ($\mu$=210) and the AAG ($\mu$=70) from the Average Grey 2 values represented in schematic nuclei. Estimation of the CI using SAG and AAG values and the described formula. The rectangle shows the region of homology in the representation of the nucleus, with the fluorescence shared between the Sarda (Green) and the Alien (Red) genomes.

Figure 3. Sarda-probe compared to other genomes over Sarda breed chromosomes. Sarda-probe showing red fluorescence was compared to green-probes of Sarda (a), Maltese (b), Murciana (c), Sheep (d) and Cattle (e). Chromosome arms show yellow-like fluorescence in all cases (under DAPI counterstain) whereas centromeres show different red and green contributions in every image. Bars below represent final contribution of each probe to the centromeres hybridization.

# Abbreviations list

CI; W-CGH; DIA;

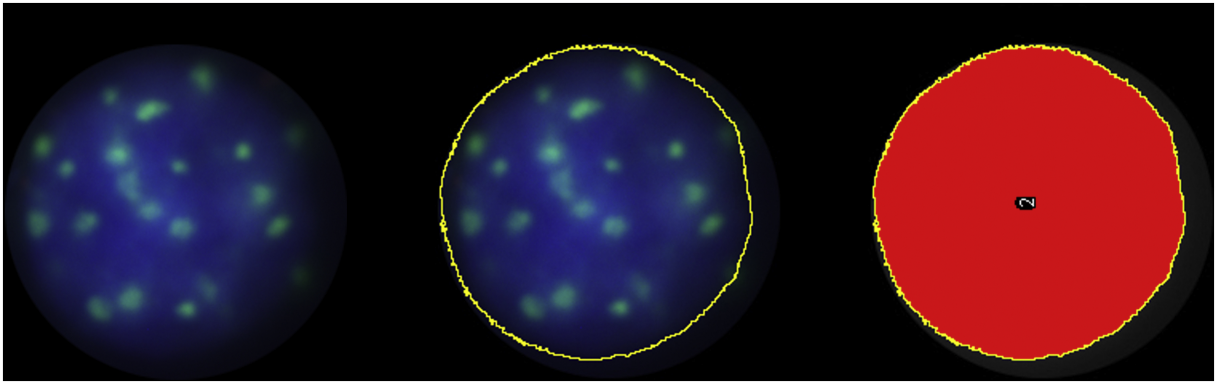

Figure 1

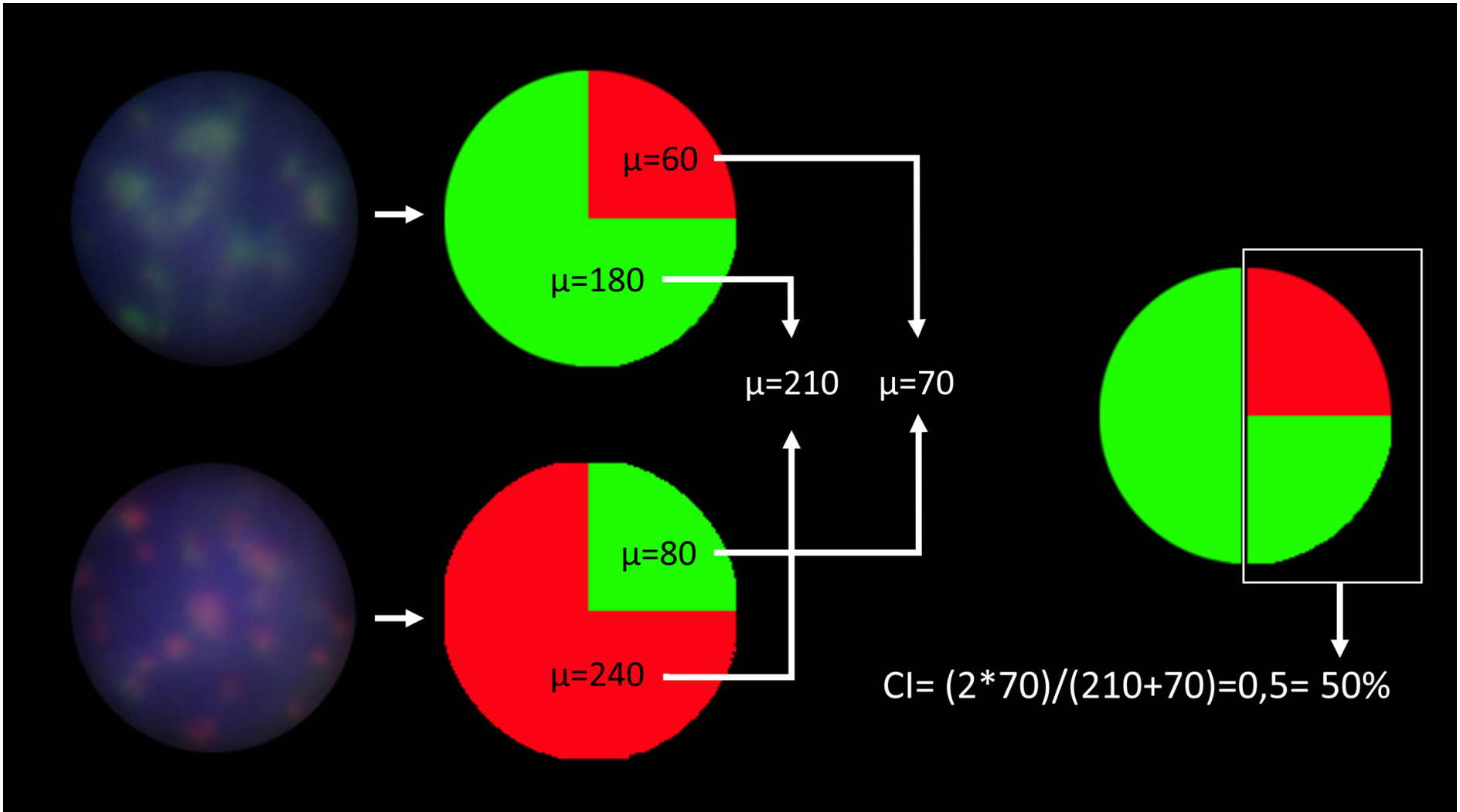


Figure 2

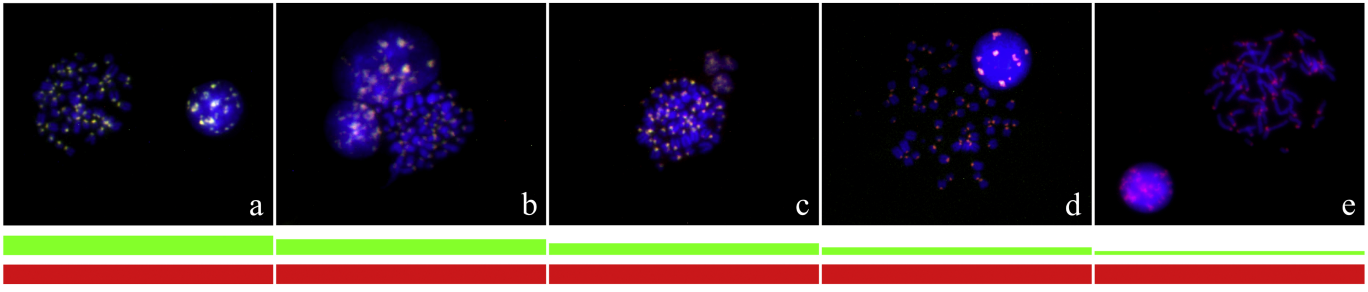


Figure 3