Short Communication

Optimization of Pre-Sampling Conditions for Salivary Genomic DNA Extraction: A Single-Donor Analysis

Sang Soo Park[1,2], Hyoungjin Choi[1,2], Jong Bhak[1,2]

1. Korean Genomics Center (KOGIC), Ulsan National Institute of Science and Technology, Ulsan, South Korea; 2. Department of Biomedical Engineering, College of Information-Bio Convergence Engineering, Ulsan National Institute of Science and Technology, Ulsan, South Korea

Objectives: Saliva-based diagnostics are expanding, yet the impact of common pre-sampling behaviors on saliva-derived genomic DNA (gDNA) recovery is not well quantified. This pilot study aimed to assess how tooth brushing, gargling, and water intake influence salivary gDNA yield and purity within a controlled single-donor setting.

Methods: Eight pre-sampling conditions were evaluated using all combinations of tooth brushing, gargling, and water intake. Saliva was collected in triplicate per condition, and cells were isolated for gDNA extraction using a Qiagen kit (Cat. No. 158023). Main effects and interactions on DNA yield and purity were analyzed by three-way ANOVA.

Results: Tooth brushing was the dominant contributor to variability in salivary DNA yield, accounting for 62.9% of the total variation, more than 11-fold greater than gargling. Water intake showed no significant effect on yield. Controlling brushing status reduced within-donor variability in DNA recovery across conditions.

Conclusions: In this single-donor pilot analysis, tooth brushing status exerted the strongest influence on salivary gDNA yield, whereas water intake had minimal impact. These results provide quantitative guidance for controlling pre-sampling behaviors to improve within-sample consistency in saliva gDNA extraction and motivate follow-up validation in larger, multi-donor cohorts.

Correspondence: papers@team.qeios.com — Qeios will forward to the authors

Sangsoo Park and Hyoungjin Choi contributed equally to this work.

1. Introduction

Saliva is a non-invasive, easily accessible biospecimen that offers substantial promise for genomic and epigenomic applications.[1] Among its components, epithelial cells serve as a valuable source of host genomic DNA (gDNA), particularly for studies aiming at methylation profiling, SNP detection, and structural variant analysis.[2][3][4][5][6][7][8] However, the quantity and quality of gDNA obtained from saliva can vary significantly depending on the donor's pre-sampling behavior, particularly in terms of gDNA yield and purity, both of which are critical factors for downstream genomics analysis. [9][10]

Despite the increasing adoption of saliva in clinical and research settings, there remains a lack of optimized protocols that account for daily habits such as brushing, gargling, or water intake prior to collection. [11][12][13] These behaviors can influence the gDNA yield and purity. For instance, food or drink consumption prior to collection can dilute salivary cell content and alter protein composition, thereby reducing DNA recovery and compromising spectrophotometric purity measurements. Similarly, mechanical stimulation from tooth brushing can increase the shedding of epithelial cells but may also introduce extraneous materials that interfere with extraction efficiency. [9][10][14]

This study aims to establish an optimized gDNA extraction protocol by systematically comparing eight distinct pre-sampling conditions, each defined by a unique combination of brushing, gargling, and water intake. Through triplicate sampling and quantitative analysis, we seek to identify the pre-sampling condition that maximizes gDNA yield while maintaining acceptable purity from saliva-derived cells, with the aim of providing reproducible and pilot evidence to guide pre-sampling control.

2. Materials and Methods

2.1. Study design and pre-sampling conditions

Saliva samples were collected under eight pre-sampling conditions defined by three factors: tooth brushing (+/–), gargling (+/–), and water intake (+/–). Each condition was sampled in triplicate (n = 3 per condition). All samples were provided by a single healthy adult volunteer. For the brushing condition, the donor brushed their teeth for 3 min using a standard toothbrush. For the mouthwash gargling condition, the donor gargled with 20 mL of Listerine® Cool Mint mouthwash for 30 s, then expectorated without additional water rinsing, following the product usage guidance (20 mL for 30 s). For the water intake

condition, the donor drank 500 mL of water. To control for timing-related variability, saliva collection was initiated 1 hour after completion of the assigned pre-sampling procedures for all conditions.

2.2. Saliva collection and handling

Whole saliva (5 mL) was collected into a 50 mL conical tube for each condition. All centrifugation steps were performed at room temperature (RT). Wide-bore pipette tips were used for resuspension and transfers to minimize mechanical shear.

2.3. Isolation and washing of saliva-derived cells

To isolate saliva-derived cells, 6 volumes of 1× PBS were added to 5 mL of saliva, followed by gentle mixing by pipetting and tube rotation at RT for 5 min. Samples were centrifuged at 500 × g for 5 min, and the supernatant was discarded. The pellet was resuspended in 5 mL of 1× PBS and washed again by repeating the same mixing and centrifugation step (500 × g, 5 min). The suspension was centrifuged at 500 × g for 5 min, the pellet was resuspended in 1 mL of 1× PBS, and the suspension was aliquoted into four 1.5 mL microcentrifuge tubes using wide-bore tips. Tubes were centrifuged at 2,000 × g for 10 min to pellet cells prior to genomic DNA extraction.

2.4. Genomic DNA extraction

Genomic DNA (gDNA) was extracted from saliva-derived cell pellets using the Qiagen Puregene Blood Core Kit (Cat no. 158023) according to the manufacturer's protocol with the following steps.

1. **Cell lysis.** Pellets were lysed with 275 µL of Cell Lysis Solution and 1.5 µL of Proteinase K, mixed by inversion (25 times), and incubated at 55°C for 1 h.
2. **RNA removal.** 1.5 µL of RNase A Solution was added, mixed by inversion (25 times), and incubated at 37°C for 15 min.
3. **Protein precipitation.** 100 µL of Protein Precipitation Solution was added, vortexed for 20 s, incubated on ice for 5 min, and centrifuged at 16,000 × g for 3 min.
4. **DNA precipitation.** The supernatant was transferred to a new tube containing 300 µL of isopropanol, mixed by gentle inversion (50 times), incubated at RT (15-25°C) for 5 min, and centrifuged at 16,000 × g for 5 min.
5. **DNA wash.** The pellet was washed with 300 µL of 70% ethanol, centrifuged at 16,000 × g for 1 min, and the ethanol was removed. The pellet was air-dried for ~5 min (avoiding over-drying).

6. **DNA hydration.** DNA was hydrated in 100 µL of DNA Hydration Solution, vortexed for 5 s at medium speed, and incubated at 65°C for 1 h to fully dissolve the DNA.

2.5. DNA quantification and purity assessment

DNA concentration was measured using both Qubit fluorometric quantification and a NanoDrop spectrophotometer. For fluorometric quantification, DNA concentration was determined using the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. NanoDrop absorbance at 260 nm (A260) was used for spectrophotometric concentration estimation, and purity was evaluated using the A260/A280 ratio. Total DNA yield (ng) was calculated from the measured concentration and elution volume (100 µL).

2.6. Statistical analysis

All analyses were performed on triplicate measurements per condition (n = 3 per condition). Total DNA yield and A260/A280 values were summarized as mean ± SD. The coefficient of variation (CV, %) for total DNA yield in each condition was calculated as (SD/mean) × 100. To identify which pre-sampling factor most strongly influences salivary DNA yield, a three-way ANOVA was performed using a 2^3 full-factorial design (brushing × gargling × water intake). Qubit fluorometric data were used as the primary outcome due to their specificity for double-stranded DNA. Effect sizes were reported as the percentage of total variation explained by each factor. A p-value < 0.05 was considered statistically significant. All statistical analyses and visualizations were performed using GraphPad Prism 11.0.0.

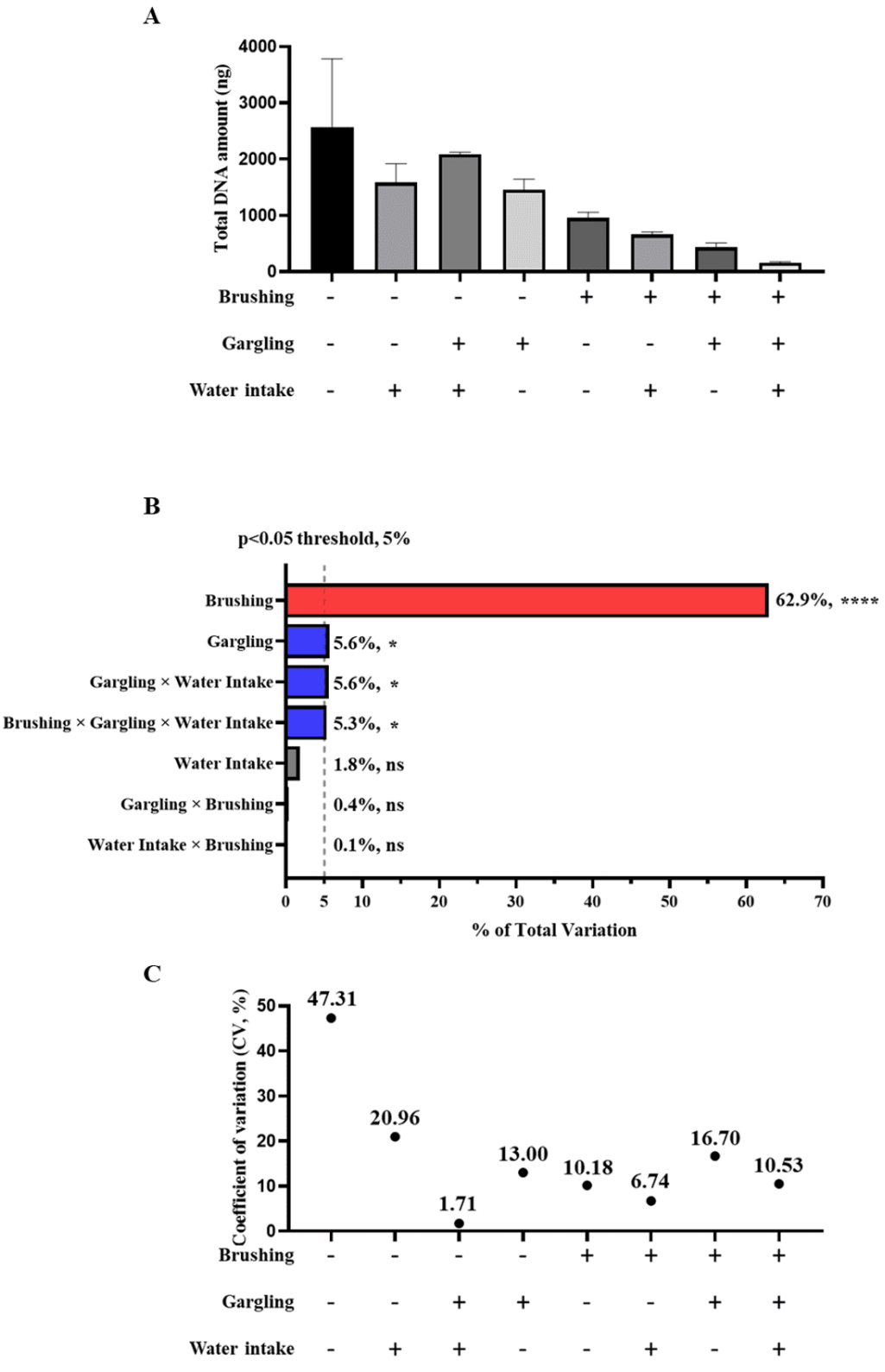
2.7. Ethics statement

This study was approved by the Institutional Review Board of Ulsan National Institute of Science and Technology (UNIST IRB; Approval No. UNISTIRB-15-19-A). All procedures were conducted in accordance with relevant institutional guidelines. Written informed consent was obtained from the participant prior to sample collection.

3. Results

3.1. Tooth Brushing Significantly Reduces DNA Yield While Maintaining Quality

doi.org/10.32388/OHJC5X



doi.org/10.32388/OHJC5X

Figure 1. Effects of pre-collection behaviors on saliva gDNA yield and reproducibility. (A) Total DNA amount (ng) across eight conditions combining brushing (+/–), gargling (+/–), and water intake (+/–). Bars show mean ± SD (n=3). **(B)** Three-way ANOVA partitioning of variance in DNA yield. Brushing accounted for the largest proportion of total variation (62.9%, ****), while gargling and selected interaction terms contributed smaller fractions (~5.6%, *). Water intake alone was not significant (1.8%, ns). *p<0.05; ns, not significant. **(C)** Reproducibility expressed as CV%. The no-intervention group (-/-/-) showed the highest variability (CV 47.31%), whereas the brushing-/gargling+/water+ condition showed the lowest variability (CV 1.71%).

Across the eight pre-collection conditions, DNA yield consistently differed by brushing status; non-brushing conditions produced higher yields, whereas brushing-containing conditions showed reduced DNA yield (Fig. 1A). To identify the main factor of the yield differences, we used a three-way ANOVA and calculated how much of the total yield variation was explained by each factor and their interactions. Each percentage indicates how much a given factor contributes to the DNA-yield differences observed across the eight conditions. Brushing accounted for 62.9% of the total yield variation, meaning that changes in brushing status (brushing vs. no brushing) explain 62.9% of the yield differences observed across the eight conditions. Thus, brushing is the single most influential factor affecting DNA recovery under the tested pre-collection behaviors (Fig. 1B, Table S1). Reproducibility analysis showed the highest variability in the no-intervention condition (CV 47.31%) and the lowest variability in the brushing-/gargling+/water+ condition (CV 1.71%), indicating the most consistent recovery under that setting (Fig. 1C, Table S2).

3.2. Pre-sampling conditions do not affect DNA purity

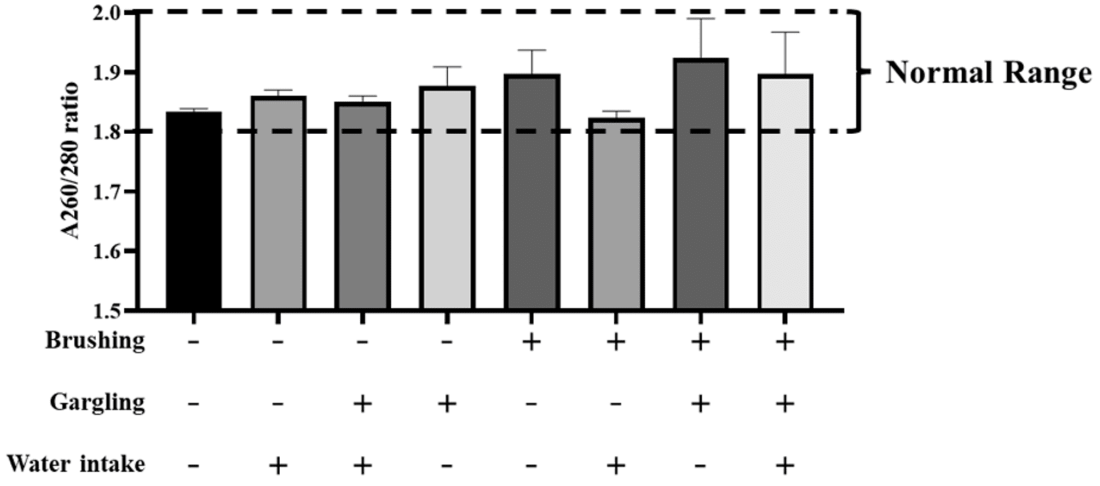


Figure 2. DNA purity across pre-sampling conditions assessed by the A260/280 ratio. The A260/280 absorbance ratio of saliva-derived genomic DNA (gDNA) was measured across pre-sampling conditions (each +/−; eight conditions). Bars show mean values and error bars indicate standard deviation (SD) from replicate measurements (n = 3). Horizontal lines in the figure indicate the acceptable purity range (1.8-2.0).

DNA purity was evaluated using the A260/280 ratio (Figure 2). Across all conditions, ratios remained within the acceptable range (1.8-2.0), and no condition-specific deviations were observed. These results indicate that pre-sampling interventions primarily affect the amount and variability of recovered DNA, while measured DNA purity remains stable across conditions.

4. Discussion

While this study provides a foundational protocol for saliva-based gDNA extraction, several important limitations must be acknowledged.

First, all samples were derived from a single healthy donor, which limits the generalizability of our findings. Inter-individual variability in oral epithelial composition, saliva viscosity, and microbiome content could substantially influence DNA yield and extraction efficiency. For example, differences in the proportion of oral epithelial cells versus immune cells (e.g., neutrophils and lymphocytes) in saliva can affect the relative contribution of host gDNA versus microbial or immune-cell-derived DNA. Individuals with higher oral microbial loads may yield lower-purity gDNA due to microbial DNA contamination,

while those with altered epithelial turnover rates may exhibit different cell yields per unit volume of saliva. These biological differences are not captured by single-donor studies and may lead to protocol recommendations that perform well only under specific donor conditions. Thus, the optimal condition identified here may not translate directly to broader populations.

Second, each pre-sampling condition was tested on separate days, rather than simultaneously or in a randomized order. This introduces the possibility of day-to-day biological fluctuations or environmental factors confounding the effects attributed to brushing, gargling, or water intake. A crossover or block-randomized design would improve the robustness of future studies.

Third, the study did not include individuals with oral conditions such as gingivitis, mucosal inflammation, or xerostomia, which are known to affect epithelial cell shedding and saliva composition. Therefore, the applicability of our standardized protocol to clinical populations remains to be validated.

Notably, the non-brushing condition with gargling and water intake (without brushing) showed the best reproducibility. One possible reason is that gargling may help reduce contaminating materials in the mouth before collection, and water intake may make it easier to produce saliva, leading to more consistent sampling.

Collectively, while our results offer a useful starting point for optimizing saliva gDNA protocols, further validation across diverse donors, health conditions, and sampling schemes is essential to establish a broadly applicable standard.

5. Conclusion

We recommend water intake and gargling (without brushing) as the optimal pre-sampling condition for saliva-based genomic DNA extraction, balancing moderate yield with the best experimental reproducibility. This study demonstrates that an effective protocol requires considering reproducibility and feasibility alongside yield optimization, providing a reliable methodological foundation for genomic studies utilizing saliva samples.

Supplementary Material

Source of Variation	SS	df	F (df_n, df_x)	p-value	% of Total Variation
Brushing	11,281,959	1	F (1, 16) = 54.94	< 0.0001 ****	62.9%
Gargling	1,010,651	1	F (1, 16) = 4.92	0.041 *	5.6%
Water Intake	321,322	1	F (1, 16) = 1.57	0.229 ns	1.8%
Water Intake × Gargling	995,115	1	F (1, 16) = 4.85	0.043 *	5.6%
Water Intake × Brushing	17,658	1	F (1, 16) = 0.086	0.773 ns	0.1%
Gargling × Brushing	63,963	1	F (1, 16) = 0.311	0.585 ns	0.4%
Water × Gargling × Brushing	954,807	1	F (1, 16) = 4.65	0.047 *	5.3%
Residual	3,285,824	16			

Table S1. Three-way ANOVA summary for salivary total DNA yield (Qubit). Statistical analysis was performed using a 2^3 full-factorial design (Brushing × Gargling × Water Intake) with Type III sums of squares (n = 3 per condition, df*residual* = 16). ****p < 0.0001; *p < 0.05; ns, not significant.

doi.org/10.32388/OHJC5X

Brushing	Gargling	Water Intake	Rep 1 (ng)	Rep 2 (ng)	Rep 3 (ng)	Mean (ng)	CV (%)
–	–	–	2,620	1,330	3,760	2,570	47.31
–	–	+	1,710	1,840	1,210	1,587	20.96
–	+	–	1,240	1,540	1,590	1,457	12.99
–	+	+	2,100	2,045	2,112	2,086	1.71
+	–	–	866	1,060	946	957	10.18
+	–	+	624	654	712	663	6.74
+	+	–	494	458	354	435	16.70
+	+	+	177	146	151	158	10.53

Table S2. Coefficient of variation (CV%) of salivary total DNA yield (Qubit) across all pre-sampling conditions. Data represent individual replicate values and mean ± CV% from triplicate measurements (n = 3 per condition). Conditions are defined by tooth brushing, gargling, and water intake status (–, not performed; +, performed).

Abbreviations: Rep, replicate; CV, coefficient of variation.

Statements and Declarations

Funding

This work was supported by the Research Project Funded by the Ulsan City Research Fund (Grant No. 1.200047.01) of Ulsan National Institute of Science and Technology (UNIST).

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author Contributions

- Sangsoo Park: Investigation, Writing – original draft, Writing – review & editing.

- Hyoungjin Choi: Data curation, Formal analysis, Writing – review & editing.
- Jong Bhak: Conceptualization, Methodology, Supervision, Writing – review & editing.

Data Availability

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

References

1. ^*Bonne NJ, Wong DT (2012). "Salivary Biomarker Development Using Genomic, Proteomic And Metabolomic Approaches." Genome Med. **4**:82.*
2. ^*Theda C, Hwang SH, Czajko A, Loke YJ, Leong P, Craig JM (2018). "Quantitation Of The Cellular Content Of Saliva And Buccal Swab Samples." Sci Rep. **8**:6944.*
3. ^*Dawes C (2003). "Estimates, From Salivary Analyses, Of The Turnover Time Of The Oral Mucosal Epithelium In Humans And The Number Of Bacteria In An Edentulous Mouth." Arch Oral Biol. **48**:329–336.*
4. ^*Bruinsma FJ, Joo JE, Wong EM, Giles GG, Southey MC (2018). "The Utility Of DNA Extracted From Saliva For Genome-Wide Molecular Research Platforms." BMC Res Notes. **11**:8.*
5. ^*Abraham JE, Maranian MJ, Spiteri I, Russell R, Ingle S, Luccarini C, Earl HM, Pharoah PP, Dunning AM, Caldas C (2012). "Saliva Samples Are A Viable Alternative To Blood Samples As A Source Of DNA For High Throughput Genotyping." BMC Med Genomics. **5**:19.*
6. ^*Gudiseva HV, Hansen M, Gutierrez L, Collins DW, He J, Verkuil LD, Danford ID, Sagaser A, Bowman AS, Salowe R, Sankar PS, Miller-Ellis E, Lehman A, O'Brien JM (2016). "Saliva DNA Quality And Genotyping Efficiency In A Predominantly Elderly Population." BMC Med Genomics. **9**:17.*
7. ^*Bahlo M, Stankovich J, Danoy P, Hickey PF, Taylor BV, Browning SR, Australian, C. New Zealand Multiple Sclerosis Genetics, Brown MA, Rubio JP (2010). "Saliva-Derived DNA Performs Well In Large-Scale, High-Density Single-Nucleotide Polymorphism Microarray Studies." Cancer Epidemiol Biomarkers Prev. **19**:794–798.*
8. ^*Samson CA, Whitford W, Snell RG, Jacobsen JC, Lehnert K (2020). "Contaminating DNA In Human Saliva Alters The Detection Of Variants From Whole Genome Sequencing." Sci Rep. **10**:19255.*
9. a, b*Hughes SR, Chapleau RR (2019). "Comparing DNA Quantity And Quality Using Saliva Collection Following Food And Beverage Consumption." BMC Res Notes. **12**:165.*

10. ^a, ^b*Rogers NL, Cole SA, Lan HC, Crossa A, Demerath EW (2007). "New Saliva DNA Collection Method Compared To Buccal Cell Collection Techniques For Epidemiological Studies." Am J Hum Biol. **19**:319–326.*
11. ^*Lim Y, Totsika M, Morrison M, Punyadeera C (2017). "The Saliva Microbiome Profiles Are Minimally Affected By Collection Method Or DNA Extraction Protocols." Sci Rep. **7**:8523.*
12. ^*Nunes LA, Mussavira S, Bindhu OS (2015). "Clinical And Diagnostic Utility Of Saliva As A Non-Invasive Diagnostic Fluid: A Systematic Review." Biochem Med (Zagreb). **25**:177–192.*
13. ^*Kuchler EC, Tannure PN, Falagan-Lotsch P, Lopes TS, Granjeiro JM, Amorim LM (2012). "Buccal Cells DNA Extraction To Obtain High Quality Human Genomic DNA Suitable For Polymorphism Genotyping By PCR-RFLP And Real-Time PCR." J Appl Oral Sci. **20**:467–471.*
14. ^*Mendoza AC, Volante BB, Hernandez ME, Mendoza CC, Pliego AF, Gonzalez HA, Juarez HE (2016). "Design Of A Protocol For Obtaining Genomic DNA From Saliva Using Mouthwash: Samples Taken From Patients With Periodontal Disease." J Oral Biol Craniofac Res. **6**:129–134.*

Declarations

Funding: This work was supported by the Research Project Funded by the Ulsan City Research Fund (Grant No. 1.200047.01) of Ulsan National Institute of Science and Technology (UNIST).

Potential competing interests: No potential competing interests to declare.